RIODICAL ROOM
NERAL LIBRARY
NIV. OF MICH.

NOV 10 1937

# The Journal of Experimental Education

A periodical report of scientific investigations relating to child development, curriculum, learning, teaching, supervision, measurements, statistics, and experimental techniques.

Volume VI | September, 1937 | Number 1

## LEARNING, TEACHING, SUPERVISION

CONTENTS

PAGE

A Comparative Experimental Study of the Pupil-Self-Study Method and the Modern-Systematic Method of Teaching Spelling: *C. W. Dupee* ..... 1

The Relative Efficiency of the "Separate" and "Together" Methods of Teaching Homonyms: *Keith C. Harder* ..... 7

A Comparison of Four Types of Spelling Tests for Diagnostic Purposes: *Joe E. Moore* ..... 24

A Study of Status and Improvement of College Freshmen in Certain Skills of English Composition: *Wilma Leslie Garnett* ..... 29

The Evaluation of a Technique for Measuring Improvement in Oral Composition: *R. F. Netzer* ..... 35

The Relationship Between Scores on the Scholastic Aptitude Test and College Marks in Chemistry: *M. Richard Dickter* ..... 40

A Study of Aural Learning With and Without the Speaker Present: *J. Edwin Loder* ..... 46

Retroactive Inhibition as a Function of Age, Intelligence, and the Duration of the Interpolated Activity: *Sister M. Florence Louise Lahey* ..... 61

The Conditioned Reflex Technique Applied to a Less Specialized Type of Learning: *Merrill T. Eaton* ..... 68

Variations of Normal Children: *Humbert G. Stout* ..... 84

Solving Problems of Action: A Proposed Procedure for the Solution of Practical Problems: *P. A. Maxwell* ..... 101

Methods of Teaching Applied to Junior Business Training in Philadelphia: *Harold B. Buckley* ..... 106

$4.00 a Year | Published Quarterly | $1.50 a Copy

Democrat Printing Company | Madison, Wisconsin

# EDITORIAL BOARD

## JOURNAL OF EXPERIMENTAL EDUCATION


**A. S. Barr, Chairman, Professor of Education, University of Wisconsin, Madison, Wis.**

Carter V. Good, Professor of Education, University of Cincinnati, Cincinnati, Ohio. Editorially responsible for materials on learning, teaching and supervision, published each September.

J. Paul Leonard, Associate Professor of Education, Stanford University, Stanford, California. Visiting Professor, Atlanta University. Editorially responsible for materials on curriculum construction, published each June.

Edward E. Cureton, Associate Professor of Education, Alabama Polytechnic Institute, Auburn, Alabama. Editorially responsible for materials on measurements, statistics, and methods of experimental research, published each March.

George D. Stoddard, Director, Child Welfare Research Station, State University of Iowa, Iowa City, Iowa. Editorially responsible for materials on child welfare, guidance, and development, published each December.

## CONTRIBUTING EDITORS

William A. Brownell, Professor of Educational Psychology, Duke University, Durham, North Carolina.

Leo J. Brueckner, Professor of Education, University of Minnesota, Minneapolis, Minnesota.

Oscar K. Buros, Professor of Education, Rutgers University, New Brunswick, New Jersey.

Otis W. Caldwell, General Secretary, The American Association for the Advancement of Science, Boyce Thompson Institute for Plant Research, Yonkers, New York.

H. L. Caswell, Professor of Education, Teachers College, Columbia University, New York City.

Ellsworth Collings, Dean, College of Education, University of Oklahoma, Norman, Oklahoma.

Herbert S. Conrad, Assistant Professor of Education, University of California, Berkeley, California.

Stephen M. Corey, Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.

Robert A. Davis, Professor of Education, University of Colorado, Boulder, Colorado.

Harl R. Douglass, Professor of Education, University of Minnesota, Minneapolis, Minnesota.

Jack W. Dunlap, Associate Professor of Education, Fordham University, New York City; University of Rochester, Rochester, New York.

Frank S. Freeman, Professor of Education, Graduate School of Education, Cornell University, Ithaca, New York.

Paul H. Furfey, Acting Head of Department of Sociology, Catholic University of America, Washington, D. C.

M. E. Herriott, Principal, Central Junior High School, Los Angeles, California.

L. Thomas Hopkins, Curriculum Specialist, Lincoln School of Teachers College, 425 West 123rd Street, New York City.

Kai Jensen, Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.

Harold E. Jones, Professor of Psychology, Director of Research, Institute of Child Welfare, University of California, Berkeley, California.

Victor H. Kelley, Assistant Director of Research and Guidance, Phoenix Union High School, Phoenix, Arizona.

Edward A. Lincoln, Consulting Psychologist, Halifax, Massachusetts.

T. E. Newland, Pennsylvania State Department of Education, Chief of Special Education, Harrisburg, Pennsylvania.

Victor H. Noll, Professor of Education, Rhode Island State College, Kingston, Rhode Island.

Lois Hayden Meek, Professor of Education, Director Individual Development Guidance, Teachers College, Columbia University, New York City.

C. W. Odell, Associate Professor of Education, University of Illinois, Urbana, Illinois.

Willard C. Olson, Professor of Education, Director of Research in Child Development, University of Michigan, Ann Arbor, Michigan.

W. E. Peik, Professor of Education, University of Minnesota, Minneapolis, Minnesota.

S. L. Pressey, Professor of Educational Psychology, Ohio State University, Columbus, Ohio.

Paul T. Rankin, Supervising Director, Research and Information Service, Board of Education, Detroit, Michigan.

Clarence E. Ragsdale, Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.

H. H. Remmers, Director, Division of Educational Reference, Professor of Education and Psychology, Purdue University, Lafayette, Indiana.

Robert T. Rock, Jr., Head, Dept. of Educational Psychology, Graduate School, Fordham University, 233 Broadway, New York, New York.

G. M. Ruch, Editorial Department, Scott, Foresman and Company, Chicago, Illinois.

P. J. Rulon, Assistant Professor of Education, Graduate School of Education, Harvard University, Cambridge, Massachusetts.

Peter Sandiford, Professor of Educational Psychology, Director, Educational Research, University of Toronto, Toronto, Canada.

Douglas E. Scates, Director, School Research, Cincinnati Public Schools, Cincinnati, Ohio.

V. M. Sims, Associate Professor of Psychology, University of Alabama, University, Alabama.

Paul W. Terry, Professor of Educational Psychology, University of Alabama, University, Alabama.

Helen Thompson, Research Associate, Yale University, New Haven, Connecticut.

Herbert A. Toops, Professor of Psychology, Statistics and College Personnel, Ohio State University, Columbus, Ohio.

T. L. Torgerson, Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.

Helen M. Walker, Associate Professor of Education, Teachers College, Columbia University, New York, New York.

John N. Washburne, Professor of Education, Syracuse University, Syracuse, New York.

Beth L. Wellman, Research Associate Professor, Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.

Paul A. Witty, Professor of Education, Northwestern University, Evanston, Illinois.

Journal of Experimental Education

Volume VI | September, 1937 | Number 1

# A COMPARATIVE EXPERIMENTAL STUDY OF THE PUPIL-SELF-STUDY METHOD AND THE MODERN-SYSTEMATIC METHOD OF TEACHING SPELLING*

C. W. Dupee
*State Teachers College*
*East Stroudsburg, Pennsylvania*

## The Problem

The purpose of this study was to compare by means of experimentation the pupil-self-study method and the modern-systematic method of teaching spelling with regard to: (1) the amount of learning achieved; (2) the rate at which learning is achieved; (3) the extent to which learning acquired by each of the two methods is retained; and (4) the use pupils make of the time saved through exemptions.

## Sources of the Data

Third, fourth, fifth, and sixth grade pupils and teachers in five different school systems in the northeastern part of the state of Pennsylvania furnished the data. The names and the locations of the school systems that participated are as follows:

Bangor Public Schools, Bangor, Pa.

East Stroudsburg Public Schools, East Stroudsburg, Pa.

Pen Argyl Public Schools, Pen Argyl, Pa.

Stroudsburg Public Schools, Stroudsburg, Pa.

Training School, State Teachers College, East Stroudsburg, Pa.

## Distribution of Pupils and Teachers

The distribution of the pupils and the teachers in the various grades is revealed in Table I. The table is read: The total number of third grade pupils participating in the experiment was 305; the total number of third grade teachers participating in the experiment was 9. The data for grades four, five, and six are read in the same manner as the data for grade three.

* This report is based on a dissertation which the writer submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Education of New York University.

Table I

Distribution of Pupils and Teachers by Grades

| Grade | Total Number of Pupils Participating | Total Number of Teachers Participating |
|---|---|---|
| 3 | 305 | 9 |
| 4 | 346 | 10 |
| 5 | 299 | 9 |
| 6 | 298 | 8 |
| Total | 1248 | 36 |

## Distribution of Pupils by Classes

The pupil distribution in the various classes in a given grade is indicated in Table II. The table reads as follows: The number of pupils in the third grade classes was 46, 39, 38, 37, 34, and so on. The mean number of pupils in the third grade classes was 33.9. The data for the remaining grades are read in the same manner as the data for grade three.

Table II

Distribution of Pupils by Classes

| Grade | Enrollment by Classes | Mean |
|---|---|---|
| 3 | 46 39 38 37 34 31 29 28 23 | 33.9 |
| 4 | 40 40 40 39 38 35 30 28 28 28 | 34.6 |
| 5 | 40 40 36 35 33 32 30 28 25 | 33.2 |
| 6 | 47 44 39 37 35 34 34 28 | 37.2 |

Table II shows that the classes were large rather than small and were therefore typical of most classes in city public schools. Hence the results are practical and not theoretical.

## Qualifications of Teachers

Table III records the amount of training of the teachers in the various grades. The table shows that all of the teachers had two or more years of training. Twenty-five per cent of the teachers participating had a bachelor's or a master's degree. Many of the teachers listed as graduates of a two year

normal course had earned additional credits toward the bachelor's degree. The teachers were, according to the teacher training standards of Pennsylvania, well trained.

TABLE III

AMOUNT OF TRAINING OF TEACHERS

| Grade | Graduate 2 Year Normal Course | Bachelor's Degree | Master's Degree* | Total No. of Teachers |
|---|---|---|---|---|
| 3 | 7 | 1 | 1 | 9 |
| 4 | 8 | 1 | 1 | 10 |
| 5 | 6 | 2 | 1 | 9 |
| 6 | 6 | 1 | 1 | 8 |
| Total | 27 | 5 | 4 | 36 |

The teachers' teaching experience is indicated in Table IV. The table shows that the teachers as a group were mature. The mean number of years of teaching experience was 14.97 or approximately 15 years. None of the teachers was inexperienced or teaching for the first time. Only one of the thirty-six teachers included in the study was in her second year of teaching. Only six teachers or 16 2/3 per cent of the total number of teachers had taught less than five years. Seven of the teachers, or approximately 20 per cent of the entire group, had taught for twenty-five or more years. From the standpoint of experience, the teachers were well able to teach spelling.

TABLE IV

TEACHING EXPERIENCE OF THE TEACHERS

| Grade | Years 0–.9 | Years 1–1.9 | Years 2–4.9 | Years 5–9.9 | Years 10–14.9 | Years 15–24.9 | Years 25–40 |
|---|---|---|---|---|---|---|---|
| 3 | | | 2 | 3 | | 2 | 2 |
| 4 | | | 3 | 3 | 2 | | 2 |
| 5 | | 1 | | 1 | 2 | 4 | 1 |
| 6 | | | | 3 | 1 | 2 | 2 |
| Mean | | | | | 14.97 | | |

## EQUATING PUPILS AND TEACHERS

The alternate method was used in equating such variable factors as initiative, mental capacity, attitudes, home environment, etc., on the part of both pupils and teachers. That is to say, a teacher and his or her pupils proceeded by the pupil-self-study method for a period of twenty days; during the next twenty-day period said teacher and pupils alternated methods and proceeded by the modern-systematic method.

* The teachers with the master's degree were training teachers in the Training School at the State Teachers College at East Stroudsburg, Pa. A very large portion of the teaching done in their grades was done by student-teachers under the direction of the training teachers.

As a precaution against any advantages one of the methods might have over the other because of initial use, approximately one half of the teachers in each of the four grades participating, together with their pupils, used the pupil-self-study method as the initial or beginning method. The remaining teachers and pupils used the modern-systematic method as the initial method. Thereafter the groups alternated methods with each successive twenty-day period.

## PROCEDURE

### *General*

1. A meeting of the teachers concerned was held in each of the school systems participating. At these meetings, the purpose of the experiment, the experimental procedure, and other items of interest or importance were explained. By the time the meetings were concluded, the teachers in each grade had divided themselves into two groups in such a manner that approximately one half of the total number of pupils in a given grade could use the pupil-self-study method as the initial or beginning method, while the other half of the grade used the modern-systematic method as the initial method.

2. The experiment continued for four learning periods of twenty days each. Each teacher alternated the method of teaching with each successive learning period. For example, a teacher who taught by the modern-systematic method during the first twenty-day period taught these same pupils by the pupil-self-study method during the second twenty-day period, by the modern-systematic method again during the third period, and by the pupil-self-study method during the fourth period. Teachers using the pupil-self-study method as the initial method alternated in like manner.

3. The time allotted to spelling on the daily class schedule was twenty minutes per day.

4. The one hundred words[1] or the subject matter taught in a given grade during a learning period was the same for both groups regardless of the method of teaching or learning used.

### *Pupil-Self-Study Method*

1. Teachers using the pupil-self-study method proceeded on the first day of a twen-

[1] All words were graded and arranged according to Henry Carr Pearson and Henry Suzzallo. *Essentials of Spelling*, pp. 21–34, 46–62, 77–93, 109–123. New York: American Book Co., 1919.

ty-day learning period as follows: (1) gave a pre-study test (written form) on the one hundred words in the contract for the twenty-day period; (2) collected test papers; (3) supplied each pupil with a printed copy of the one hundred words in the contract for that period; (4) advised the pupils that they were to do their own studying when the pupil-self-study method was used, that they would be tested daily (written form) on the one hundred words in the contract, and that a pupil would be exempt from spelling, except for the final test, when he or she made a perfect score on two successive tests.

2. From the second to the nineteenth day of the learning period, inclusive, the teachers used the twenty minutes allotted to spelling on the daily class schedule for the purpose of giving the pupils who were not exempt a written test on the one hundred words in the contract for that period.

3. On the twentieth or last day of the contract period, the teachers gave a final test, written form, on the one hundred words studied during the twenty-day period. All of the pupils, exempt as well as non-exempt, were required to take the final test.

4. When a pupil became exempt, he was asked to note in diary fashion the use or uses he made of the spelling period the days on which he was exempt.

5. All test papers were scored and returned to their owners at the earliest possible convenience.

6. A chart was posted on the bulletin board at the beginning of each twenty-day learning period. The number of words misspelled or the learning status of each pupil was recorded on the chart daily.

The primary duties of the teacher when the pupil-self-study method was used were group testing, analyzing results, diagnosing individual cases, and causing or allowing each pupil to work up to capacity. In other words, the pupil-self-study method made allowances and provisions for individual differences. Teachers were requested to refrain from using mass instruction when the pupil-self-study method was employed.

*Modern-Systematic Method*

1. The twenty-day learning period was divided into four weeks or periods of five days each. The one hundred word contract for that twenty-day period was likewise divided into four units of twenty-five words each. The first twenty-five words in the contract comprised the subject matter to be taught during the first week of the contract period; the second twenty-five words comprised the subject matter for the second week; and so on for the third and fourth weeks.

2. The teachers proceeded on the first day of the twenty-day period as follows: (1) gave a pre-study test*, written form, on the one hundred words in the contract for the twenty-day period; (2) collected test papers; (3) gave pupils printed copies of the first twenty-five words in the contract for the period; (4) informed the pupils that they would all study spelling together during the regular spelling period each day; none would be exempt from spelling; and, spelling could be studied at home or outside of the regular spelling period if the pupils so desired.

3. The general procedure on the first and the last day of the twenty-day period differed from the corresponding weekly procedure to the extent that the pre-study test and the final or after-study test given on these days included the entire list of one hundred words for that period, whereas the weekly tests included only the twenty-five words for the week. (The reasons for this difference are obvious.)

4. The duties of the teacher on a given day of the week, except for the first and the last day of the twenty-day period, were:

Monday, give a pre-study test on the twenty-five words to be taught during that week; use the balance of the period for the purpose of teaching the pupils the correct spelling of some of the twenty-five words in the week's assignment.

Tuesday, use the twenty minutes allotted to spelling on that day for teaching purposes. (With the modern-systematic method, each teacher was free at all times to use his favorite or most effective technique for teaching spelling.)

Wednesday, proceed in the same manner as on Tuesday.

Thursday, give an after-study test on the twenty-five words taught during the week; use the remainder of the period for reviewing or re-teaching purposes.

Friday, conduct a review lesson with the class as a whole for the purpose of fixing in the pupils' minds the troublesome or un-

* These tests were scored but, for obvious reasons, the papers were not returned to the pupils until the end of the twenty-day period.

learned words in that portion of the contract studied to date.

5. On the twentieth or last day of the contract period, a final test was given (written form) on the one hundred words in the contract for that period.

6. The teachers recorded the results of the pre-study tests on Monday and the after-study tests on Thursday of each week on a chart posted on the bulletin board for that purpose.

With the modern-systematic method, the primary duty of the teacher was methodically directing the learning of the pupils en masse.

*Retention*

Another procedure was necessary for collecting the data on retention. One month after the date of the final tests on Contract IV, a retention test was given to each of the four grades participating. Two hundred words chosen alternately, fifty words from each of the four contracts studied or one half of the total number of words taught in any one grade, comprised the retention test for a given grade.

Scoring the retention tests and tabulating the results concluded the data-collecting part of the experiment.

## Summary of Findings

The findings of the study may be summarized as follows:

1. The mean number of words learned per pupil per contract period by each of the four grades participating was greater when the pupils studied by the pupil-self-study method than when they studied by the modern-systematic method. When the pupils studied by the pupil-self-study method the least mean number of words learned per pupil per contract period, namely, 35.47 words, was learned by the fifth grade pupils; the greatest mean number of words learned per pupil per contract period, namely, 39.79 words, was learned by the third grade pupils.

When the pupils studied by the modern-systematic method the least mean number of words learned per pupil per contract period, namely, 28.26 words, was learned by the sixth grade pupils; the greatest mean number of words learned per pupil per contract period, namely, 33.06 words, was learned by the third grade pupils.

The mean number of words learned per pupil per contract period was as follows:

| | Pupil-Self-Study Method | Modern-Systematic Method |
|---|---|---|
| Grade 3 | 39.79 | 33.06 |
| Grade 4 | 35.64 | 28.59 |
| Grade 5 | 35.47 | 28.68 |
| Grade 6 | 36.31 | 28.26 |

2. The difference in means or the extent to which the mean number of words learned per pupil per contract period by the pupil-self-study method exceeded the mean number of words learned per pupil per contract period by the modern-systematic method ranged from 6.73 words or 20.36 per cent per pupil per contract period for the third grade pupils to 8.05 words or 28.49 per cent per pupil per contract period for the sixth grade pupils.

The difference in the means for the various grades was as follows:

| | |
|---|---|
| Grade 3 | 6.73 words or 20.36 per cent |
| Grade 4 | 7.05 words or 24.66 per cent |
| Grade 5 | 6.79 words or 23.68 per cent |
| Grade 6 | 8.05 words or 28.49 per cent |

3. The mean rate of learning per pupil twenty-minute period of school time devoted to spelling was greater for each of the four grades participating when the pupils studied by the pupil-self-study method than it was when they studied by the modern-systematic method. When the pupils studied by the pupil-self-study method, the least mean number of words learned per pupil twenty-minute period of school time devoted to spelling, namely, 2.391 words, was credited to the fourth grade pupils; the highest mean rate for the pupil-self-study method, namely, 2.817 words, was credited to the sixth grade pupils.

When the pupils studied by the modern-systematic method, the least mean number of words learned per pupil twenty-minute period of school time devoted to spelling, namely, 1.395 words, was credited to the sixth grade pupils; the highest mean rate for the modern-systematic method, namely, 1.630 words, was credited to the third grade pupils.

The difference in the means for the various grades ranged from .896 of a word or 54.97 per cent for the third grade pupils to 1.422 words or 101.94 per cent for the sixth grade pupils.

The mean rate of learning per pupil twenty-minute period of school time devoted to spelling was as follows:

| | Pupil-Self-Study Method | Modern-Systematic Method |
|---|---|---|
| Grade 3 | 2.526 words | 1.630 words |
| Grade 4 | 2.391 words | 1.396 words |
| Grade 5 | 2.622 words | 1.412 words |
| Grade 6 | 2.817 words | 1.395 words |

4. In grades three, four, and six, the mean per cent of learning retained per pupil per contract period was somewhat greater when the pupils studied by the modern-systematic method than it was when they studied by the pupil-self-study method. In the fifth grade the order was reversed.

The mean per cent of learning retained per pupil per contract period when the pupils studied by the pupil-self-study method ranged from 72.82 per cent for the fifth grade pupils to 92.41 per cent for the third grade pupils; when the pupils studied by the modern-systematic method the range was from 70.64 per cent for the fifth grade pupils to 97.43 per cent for the third grade pupils.

The third grade pupils showed the greatest difference, namely, 5.02 per cent, in the mean per cent of learning retained; the fourth grade pupils showed the least difference, namely, .59 of one per cent, in the mean per cent of learning retained.

The mean per cent of learning retained by each of the four grades participating was as follows:

| | Pupil-Self-Study Method | Modern-Systematic Method |
|---|---|---|
| Grade 3 | 92.41 per cent | 97.43 per cent |
| Grade 4 | 84.40 per cent | 84.99 per cent |
| Grade 5 | 72.82 per cent | 70.64 per cent |
| Grade 6 | 75.35 per cent | 78.31 per cent |

5. The mean number of words forgotten* per pupil per contract period was somewhat greater for each of the four grades participating when the pupils studied by the pupil-self-study method than when they studied by the modern-systematic method. The difference in the means for a given grade ranged from 1.14 words for the fifth grade pupils to 2.27 words for the sixth grade pupils.

The mean number of words forgotten per pupil per contract period for each of the two methods of study employed was as follows:

| | Pupil-Self-Study Method | Modern-Systematic Method |
|---|---|---|
| Grade 3 | 3.34 words | 1.49 words |
| Grade 4 | 5.74 words | 4.02 words |
| Grade 5 | 9.41 words | 8.27 words |
| Grade 6 | 8.22 words | 5.95 words |

* The reader will please note that the terms "mean number of words forgotten" and "mean per cent of learning retained" are not synonymous in meaning.

6. The total number of twenty-minute spelling periods saved through exemptions when the pupils studied by the pupil-self-study method was 11,845. In other words, during the half of the experiment in which the pupils studied by the pupil-self-study method, the pupils saved a total of 3,948 and 1/3 clock hours of school time.

The time saved through exemptions was equal to 29.20 per cent of the total time allotted to the pupil-self-study method.

The total number of twenty-minute spelling periods the pupils in a given grade were exempt when they studied by the pupil-self-study method was as follows:

| | |
|---|---|
| Grade 3 | 2013 |
| Grade 4 | 2871 |
| Grade 5 | 3348 |
| Grade 6 | 3613 |

7. In keeping with tradition and common practice, the pupils were not permitted to save time through exemptions when they studied by the modern-systematic method.

8. Three items, namely, "did home work", "read a library book", and "drew pictures", included 92.43 per cent of the total number of entries made in the diaries which the pupils kept when they were exempt. "Did home work", meaning, prepared next day's assignment in arithmetic, geography, etc., included 38.84 per cent of the total number of entries; "read a library book" included another 38.46 per cent; and, "drew pictures" included another 15.13 per cent. The remaining 7.57 per cent of the total number of entries were sparsely distributed over a large number of items.

## Conclusions

The conclusions or answers to the problems proposed at the beginning of the study may, in accordance with the data and their treatment, be stated as follows:

1. There is practically complete certainty, more than 99.9 chances[1] in 100, that the true difference in the mean number of words learned per pupil by third, fourth, fifth, and sixth grade pupils is greater than zero when the pupils study by the pupil-self-study method and by the modern-systematic method. The difference is in favor of the pupil-self-study method.

2. The mean rate of learning per pupil twenty-minute period of school time devoted

[1] Henry E. Garrett, *Statistics in Psychology and Education*, p. 134. New York: Longmans, Green and Company, 1926.

to spelling is greater when third, fourth, fifth, and sixth grade pupils study by the pupil-self-study method than it is when they study by the modern-systematic method.

3. The mean per cent of learning retained by third, fourth, fifth, and sixth grade pupils is about the same when the pupils study by the pupil-self-study method as it is when they study by the modern-systematic method.

4. In a very large majority of cases, third, fourth, fifth, and sixth grade pupils make worthwhile use of the time they save through exemptions. The large amount of time saved, together with the use pupils make of their exemption periods, is conclusive evidence that the pupil-self-study method has great possibilities as a way or a means of providing for individual differences and of varying and enriching the curriculum.

# THE RELATIVE EFFICIENCY OF THE "SEPARATE" AND "TOGETHER" METHODS OF TEACHING HOMONYMS[1]

KEITH C. HARDER
*Chattanooga, Tennessee*

The spelling of homonyms has ever been a source of difficulty for both teachers and pupils. An examination of thirty-eight spelling books published since 1826 revealed many ingenious arrangements for grouping, defining, and illustrating homonyms in lessons.

The problem of confusions in spelling was noticed by Burnham[2] as early as 1906 when he wrote:

"In regard to certain words many persons are the victims of chronic interference of association. Examples are the noun *advice* and the verb *advise;* 'principal', an individual, and 'principle', a rule of conduct . . .

"In some respects prevalent methods might be revised with advantage. The usual method of teaching spelling of homonyms, for example, is to present them simultaneously. The result is apt to be interference of association. To avoid this they should be taught separately. The very similarities that are supposed to aid the pupil become the source of confusion. In such cases the slight difference, not the general similarity, is the important thing."

In 1911 Pearson[3] made the first experimental attack on this problem when he studied the teaching of homonyms at the Horace Mann School in New York. Each of the third, fourth, fifth, sixth, and seventh grades used had two sections with a teacher for each. The general plan of the experiment was: (1) to test the ability of each grade to spell the selected homonyms, (2) to have these homonyms taught by the regular teacher of one section of a grade by the "separate" method and by the other teacher of that grade by the "together" method, (3) to test the pupils' ability to spell these homonyms, and (4) to compare the results of the two methods. In one section of a grade the pair of homonyms was presented adjacent to each other in the same lesson, and in the other section with an interval of three or four days between the words of each pair.

The results showed: (1) a gain for the "together" method in every grade except the fifth, (2) an advantage for the "together" method for the poor spellers. The experiment was repeated at Montclair, New Jersey, with the results confirming those found by Pearson.

Several weaknesses may be observed in the study as carried out:

1. No attempt was made to control the factor of individual teacher personality.

2. The interval between the appearance of the members of a pair (three or four days) when taught by the "separate" method was too short.

3. The sections used were not equated on spelling ability.

4. A longer interval than ten days after study should elapse before testing for permanent recall.

5. The number of cases used was too small to be a representative sample.

Several of the criticisms noted above were observed by Finkenbinder[4] in his review of this study. In 1923 Finkenbinder made the second experimental attack on this problem using three fourth grade classes in the public schools of Worcester, Massachusetts. Sixty homonyms were selected, twenty-four pairs of which were from the spelling books in use, and twelve homonyms, not pairs, from outside the spelling books. These homonyms and eighty-four other words were grouped into lessons of eighteen words each, and taught during a period of eight days, ten minutes of study per day. In order to equalize the effect of the personality of each teacher on the two methods and to insure that each pupil studied both lists of words, a rotation pro-

[1] Abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Department of Education, in the Graduate College of the State University of Iowa, August, 1937.

[2] Wm. H. Burnham, "The Hygiene and Psychology of Spelling," *Pedagogical Seminary*, XIII (December, 1906), 495.

[3] Henry C. Pearson, "The Scientific Study of the Teaching of Spelling," *Journal of Educational Psychology*, II (May, 1911), 241–252.

[4] E. O. Finkenbinder, "The Spelling of Homonyms: An Experimental Investigation of Teaching Them," *Pedagogical Seminary*, XXX (September, 1923), 241–251.

cedure was employed. It is briefly described as follows:

The first twelve pairs of the forty-eight homonyms were studied by the "separate" method in one class and by the "together" in another class. The second twelve pairs were studied by the "together" method in the first class named above, and by the "separate" method in the other class. In a third class one half of the first twelve pairs and one half of the second twelve pairs were studied by the "separate" method and the remaining one half of the first and second twelve pairs were studied by the "together" method. The pupils were tested at the close of the study and in two of the three classes, one month later.

Results:

1. The preliminary test showed that in one fourth of the pairs neither word was correctly spelled; and in about one half of the pairs, only one word was misspelled.

2. In the test given immediately after study, the methods were found to be of almost equal value. On this test the middle fifty per cent of the pupils, ranked according to ability, showed a superiority for the "separate" method. The poorest spellers showed a superiority for the "together" method.

3. The results for the test one month after study show:

a. For the group of homonyms, both of which required study, the pupils taught by the "separate" method excelled in spelling accuracy those taught by the "together" method by 17.7 per cent.

b. For the group of words, only one of which required study, the "separate" method excelled the "together" method by .5 per cent.

Criticism of the study:

1. Not enough time (eight days) was given to the study as a whole.

2. In the "separate" method the interval between the studying of one homonym and the other member of a pair was too short.

3. The number of pupils (three classes) was too small to give reliable results.

4. In the analysis of the results Finkenbinder's conclusions are limited in their scope by the small numbers of pupils and homonyms. His main conclusions are based on six pairs of homonyms (one fourth of the pairs studied had both words misspelled on the pretest). It was on this group that differences large enough on which to base recommendations as to method were found. Twelve pairs were in the group which had one word spelled correctly on the pretest. This group showed an advantage of only .5 per cent for the "separate" method.

Foran[5] summarizes the present status of experimental evidence on the teaching of homonyms as follows:

"Homonyms constitute a class of words that present considerable difficulty, and the best method of teaching them cannot be defined on the basis of existing evidence. The 'together' and 'separate' methods are both used, and each may be defended on some basis or other. The early studies by Pearson support the teaching of homonyms 'together'. Later, Finkenbinder found that the 'separate' method produced better results. Neither series of experiments was entirely free from defects that would tend to invalidate the results regardless of whatever method will prove to be superior when the issue is eventually decided. Homonyms form one of the most difficult groups of words that pupils are called upon to spell. A recently published speller contains sixty-one pairs of such words, and practically all of them are commonly used. Such words as *there* and *their, eight* and *ate* occur in the early grades and are the source of much confusion."

## Object

The purpose of this investigation is to determine whether the "separate" or "together" method of presenting homonyms in lessons would result in greater spelling accuracy. In addition, the results show which method is better for each of the homonyms included in the investigation. The methods are also compared to see if either method gives better results for the highest and lowest ten per cent of the pupils.

## Definitions

Since the following terms, used in this study, have specific meanings they are defined for the reader.

*Cluster,* refers to a group of words having the same phonetic sound but differing in spelling, such as *pair-pear-pare, there-their-they're, bare-bear,* etc.

[5] Thomas George Foran, *The Psychology and Teaching of Spelling,* pp. 40–41. Washington, D. C.: The Catholic Education Press, 1934.

*Confusion,* as used in this study, means the substitution of the correct spelling of one homonym for another within the same cluster.

*Delayed recall interval,* as used in this study, refers to the period of time, eight weeks, which elapsed between the immediate recall and the delayed recall tests.

*Errors,* include both the wrong spelling of the homonym called for, and the right spelling of any other word substituted for it.

*Homonym,* as defined by Webster[6] is "a word having the same pronunciation as another, but differing from it in origin, meaning, and often, in spelling, as *bare* and *bear* . . ."

*Misspelled form,* as employed in this study, means one of the ways in which an homonym is spelled incorrectly.

*Separate,* means that the words having the same or similar sounds are presented singly, only one of a cluster appearing in a lesson.

*Right spelling,* refers only to the correct spelling of the homonym called for.

*Together,* means that the two or three words having the same or similar sounds are presented adjacent to each other in the same lesson.

## Procedure

Thirty third grade and twenty-four sixth grade classes from seven Iowa and Nebraska school systems using the *Horn–Ashbaugh Fundamentals of Spelling*[7] were used in the experiment. This text gives specific directions for a test-study method of teaching spelling.[8] The unit of instruction is the list of words assigned for one week's study. Briefly described, this method provides for a test over the words before study on Monday; a study of the words missed on this test on Tuesday; a second test over the words assigned for the week and a review test over the words taught one month previously, on Wednesday; a further study on Thursday of the words missed on the Wednesday test; and on Friday a test over all the words tested on Wednesday. The reader should note that this method provides for testing both immediately after study and review testing one month later for all the words taught.

[6] Noah Webster, *Webster's New International Dictionary of the English Language,* p. 1194. (2nd ed., unabridged). Springfield, Mass.: G. and C. Merriman Company, 1934.

[7] Ernest Horn and E. J. Ashbaugh, *The Horn–Ashbaugh Fundamentals of Spelling.* Chicago: J. B. Lippincott and Company, 1928.

[8] *Ibid.,* pp. xiii–xvi.

It was believed that the schools using this text would show less variability in teaching procedure than schools using a variety of spelling methods and textbooks. The pretest, instruction periods, immediate recall tests, and delayed recall tests covered the second semester of the school year 1934–1935.

The "separate" and "together" groups in the third grade and in the sixth grade were equated by means and standard deviations based on the pretest as follows:

TABLE I

EQUATION OF EXPERIMENTAL GROUPS

| Test | Method | No. of Cases | Arithmetic Mean | Standard Deviation |
|---|---|---|---|---|
| | | Third Grade | | |
| Immediate Recall Test | Separate | 588 | 21.112 | 11.6473 |
| | Together | 588 | 21.112 | 11.6473 |
| Delayed Recall Test | Separate | 297 | 22.0370 | 12.0560 |
| | Together | 295 | 22.0373 | 12.0615 |
| | | Sixth Grade | | |
| Immediate Recall Test | Separate | 636 | 37.5582 | 14.0856 |
| | Together | 635 | 37.5582 | 14.0856 |
| Delayed Recall Test | Separate | 279 | 36.8638 | 13.4624 |
| | Together | 279 | 36.8638 | 13.4592 |

A rotation technique provided that each teacher taught both methods and both lists of words.

## Selection of Words

The lessons were composed of homonyms and other words. The words other than homonyms were selected from the second semester lessons in the third and sixth grades of the *Horn–Ashbaugh Fundamentals of Spelling.*[9] Two hundred and nineteen homonyms were selected from Horn and Ashbaugh,[10] Ashbaugh,[11] Fitzgerald,[12] Cooper,[13] Masters,[14] Peterson,[15] and Fellows.[16] From these homonyms the sixty used in the third grade and the eighty-six used in the sixth

[9] *Ibid.*

[10] *Ibid.*

[11] E. J. Ashbaugh, *The Iowa Spelling Scales.* Bloomington, Illinois: Public School Publishing Co., 1922. Journal of Educational Research Monographs, No. 3.

[12] James A. Fitzgerald, "Words Misspelled Most Frequently by Children of the Fourth, Fifth and Sixth Grade Levels in Life Outside the School," *Journal of Educational Research,* XXVI (November, 1932), 213–218.

[13] E. J. Cooper, "The 1035 Words Most Commonly Misspelled by Minnesota Pupils in Grades Three to Eight". Master's Thesis, University of Minnesota, 1927.

[14] H. V. Masters, *A Study of Spelling Errors.* University of Iowa Studies in Education, IV, No. 4. Iowa City, Iowa, 1927.

[15] Evelyn F. Peterson, "An Analysis of Spelling Error". Master's Thesis, University of Iowa, 1930.

[16] Samuel Fallows, *A Complete Dictionary of Synonyms and Antonyms.* Chicago: F. H. Revell Company, 1883.

TABLE II

GENERAL ARRANGEMENT OF CLASSES FOR THIRD AND SIXTH GRADES ILLUSTRATING THE ROTATION TECHNIQUE

| Teacher | Test | Method | Test | Method | Test |
|---|---|---|---|---|---|
| A | Pretest | Separate List A (8 weeks) | Immediate Recall | Together List B (8 weeks) | Final |
| B | Pretest | Together List A (8 weeks) | Immediate Recall | Separate List B (8 weeks) | Final |
| C | Pretest | Together List B (8 weeks) | Immediate Recall | Separate List A (8 weeks) | Final |
| D | Pretest | Separate List B (8 weeks) | Immediate Recall | Together List A (8 weeks) | Final |

Explanatory notes for Table II

a. Every teacher taught the "separate" method for eight weeks and the "together" method for eight weeks.

b. In each eight-week period the homonyms were introduced during the first four weeks and reviewed during the next four weeks.

c. In the third grade fifteen words were taught per week; in the sixth grade twenty words were taught per week.

d. The homonyms in each grade were divided into lists A and B and were both taught in each eight-week period.

e. The immediate recall test for each grade was composed of only the homonyms which had been studied during the first eight weeks.

f. In third grade 240 words were taught, of which sixty were homonyms. In the sixth grade 320 words were taught, of which eighty-six were homonyms.

---

grade were chosen on the basis of the data available at the University of Iowa on frequency of adult usage, frequency of child usage, grade placement, and difficulty. The homonyms used in each grade were divided into two lists of approximately equal difficulty.

THIRD GRADE—LIST A

| | | | |
|---|---|---|---|
| aisle | fore | lets | to |
| I'll | four | let's | too |
| isle | hears | mail | two |
| ant | here's | male | wait |
| aunt | higher | road | weight |
| billed | hire | rode | |
| build | hours | some | |
| cents | ours | sum | |
| sense | | | |

THIRD GRADE—LIST B

| | | | |
|---|---|---|---|
| ate | its | meat | their |
| eight | it's | meet | there |
| close | knew | mete | wares |
| clothes | new | pair | wears |
| grate | knights | pare | |
| great | nights | pear | weak |
| heard | knows | right | week |
| herd | nose | write | |

SIXTH GRADE—LIST A

| | | | |
|---|---|---|---|
| accepted | baring | false | soles |
| excepted | bearing | faults | souls |
| adieu | billed | pause | to |
| ado | build | paws | too |
| affect | choir | presence | two |
| effect | quire | presents | waits |
| aisle | commence | profits | weights |
| I'll | comments | prophets | wares |
| isle | council | rooted | wears |
| assistance | counsel | routed | weave |
| assistants | consul | | we've |

SIXTH GRADE—LIST B

| | | | |
|---|---|---|---|
| addition | currants | ladies | sane |
| edition | currents | ladies' | seine |
| attendance | decent | lady's | stationary |
| attendants | descent | passed | stationery |
| canvas | fate | past | their |
| canvass | fete | plain | there |
| coarse | fathers | plane | they're |
| course | father's | praise | vary |
| companies | intents | prays | very |
| company's | intense | preys | writers |
| | | principal | writer's |
| | | principle | |

### Arrangement in Lessons

Each cluster of homonyms was placed in two booklets. The "separate" method booklet of eight lessons was so arranged that the first four lessons included homonyms grouped so that an interval of approximately two weeks elapsed between the studying of the homonyms of a cluster. The "together" method booklet of eight lessons was so arranged that the first four lessons included homonyms grouped together. In both booklets the last four lessons had no homonyms. The first four lessons appeared as review words during the second four weeks. The meaning and identification of each homonym were made clear by its use in a sentence. The words not homonyms were listed in column form.

Lesson two follows, as given in the booklets for the third grade, "together" method.

**Lesson Two**

| | | | |
|---|---|---|---|
| lace | guess | change | file |
| match | steam | engine | potatoes |
| afraid | named | hundred | |

| | | |
|---|---|---|
| herd | Mr. Jones has a fine dairy *herd*. | herd |
| heard | I *heard* the birds sing. | heard |
| ate | He *ate* an apple. | ate |
| eight | I saw *eight* birds. | eight |

Lesson two follows, as given in the booklets for the sixth grade, "separate" method.

**Lesson Two**

| | | | |
|---|---|---|---|
| slight | sweetheart | cheaper | carbon |
| occurs | finest | poultry | amounting |
| advised | moral | assure | educational |

| | | |
|---|---|---|
| isle | The *isle* was captured. | isle |
| bearing | He was *bearing* a plank. | bearing |
| assistance | She was of *assistance* to her mother. | assistance |
| adieu | His *adieu* was very formal. | adieu |
| wears | She *wears* a blue dress. | wears |
| accepted | He *accepted* the offer. | accepted |
| choir | The *choir* sings in church. | choir |
| council | The town *council* met last night. | council |

### Instructions to Teachers

Before being sent to the schools, detailed instructions were prepared and given a trial by teachers experienced in using the test-study method. The instructions included a general description of the experiment, its purpose, and procedure in so far as it affected the teacher in the classroom. Attention was called to the use of the method of teaching, and the method of study to be followed by the pupils as outlined by Horn and Ashbaugh.[16a] Detailed directions for the giving of the pretest, immediate recall, and delayed recall tests were included. In addition, each teacher received special instructions with the booklet for the method to be taught.

In teaching the "separate" method the teacher was instructed *not to* call attention to other words which have the same sound. If the pupils asked questions about other homonyms, they were to be answered but not in detail. The teacher was to keep in mind that the homonyms were not to be taught together in this part of the experiment.

In teaching the "together" method the teacher was instructed to stress the fact that here were two or three words with the *same sound but different meaning*. These two or three words were to be associated together during the learning period.

### Tests

The form of the test selected was the modified sentence containing blanks in which the pupil wrote the proper word. Since there was a possibility that pupils might guess the right form of the homonym to use in the sentence, it was decided to test each item twice, thus getting a pupil response to two situations in which the same meaning of the word was used. In making the tests and booklets, a certain meaning was selected as the one to use for the meaning of the homonym. The first sentence used in the booklets and also in the tests was called Form I. The second was called Form II, the third, Form III. The pretest was made up of Forms I and II for each list of homonyms and the immediate recall tests of Forms II and III. The delayed recall used the same forms as the pretest. Some illustrative sentences from the teachers' copy of the third grade pretest, Forms I and II, are given showing the testing of the homonym clusters *ant-aunt*, and *their-there*. The teacher was instructed to pronounce the word, read the sentence using it, and pronounce the word called for again. On the pupils' test sheets the homonyms to be spelled were omitted.

| | Item number | Sentence Used |
|---|---|---|
| | | **Form I** |
| ant | 1. | The _____ ran on the table. |
| there | 10. | I want to be _____. |
| aunt | 17. | Did you see your _____? |
| their | 24. | _____ house is red. |

[16a] Horn and Ashbaugh, *op. cit.*, pp. xiii–xvi, xxiv–xxvi.

Form II

| | | |
|---|---|---|
| aunt | 1. | I love my _____. |
| there | 10. | _____ is the flag. |
| ant | 17. | Did you see the _____ hill? |
| their | 26. | I want _____ books. |

Immediately preceding the study of the booklets, each class participating was given the pretest which included all of the homonyms that were studied in the investigation. The immediate recall test was given at the end of the eighth week and each pupil was tested on all of the homonyms he had studied during the preceding eight weeks.

At the close of the experiment the pretest was repeated. For the homonyms studied during the latter half of the experiment it was an immediate recall test and for those studied during the first half of the experiment it was a delayed recall test.

In scoring the weekly tests the teachers were instructed to mark a word wrong if it was omitted, illegibly written, if the wrong word or wrong homonym was substituted, if apostrophes were used incorrectly, if the word had been erased to correct, marked over to correct, been begun incorrectly, crossed out and rewritten, or did not have *i's* dotted or *t's* crossed.

## Method of Analyzing the Results

In the major analysis of the results the "separate" and "together" methods were compared on the basis of the number of right spellings. The statistical treatment used in reporting the main findings in this study was developed by Lindquist[17] and is called the formula for the standard error of the difference for matched groups.

$$\delta \text{ diff.} = \sqrt{\left(\frac{\delta_1^2}{N_1} + \frac{\delta_2^2}{N_2}\right)(1 - r^2)}$$

$\delta_1$ and $\delta_2$ are the standard deviations of the scores of the matched groups. $N_1$ and $N_2$ are the number of cases in each group. "$r$" is the correlation between the pretest and the test being considered.

$$\text{critical ratio} = \frac{\text{Difference in A. M.}}{\text{Standard error of a difference}}$$

The critical ratio was determined for each difference between the groups. If it was more than three, the difference was considered to be significant.

[17] E. F. Lindquist, "The Significance of a Difference Between Matched Groups." *Journal of Educational Psychology*, XXII (March, 1931), 197–204.

In determining which method was better for each homonym the comparisons were based on the number of pupils having two-right spellings.[18] A formula by Yule,[19] with the percentage tables developed by Edgerton and Paterson,[20] was used in determining whether the percentage differences in right spellings between the two methods for each homonym were significant. The formula is:

$$\delta \text{ diff. in percentage} = \sqrt{p_1^2 + p_2^2}$$

This formula is based on another by Yule[21] called the "standard error of a percentage or proportion".

$$\delta p = \sqrt{\frac{pq}{N}}$$

$p$ = The percentage or proportion
$q$ = The difference between p and 1.00
$N$ = The number of cases

## Most Frequent Errors

In addition to the above, a study of the errors was made for each of the methods of presentation in third and sixth grades. A sampling of papers was taken from the immediate and delayed recall tests, 150 from each test for each method, by selecting every third paper from all schools participating, until the desired number was obtained. The errors found on each of these test papers were tabulated and classified for each method of presentation, according to the number of misspelled forms, the most common errors with frequency of occurrence, and the total number of errors for each homonym taught.

## Results

The main results reported in the following tables show that one method is superior for immediate recall and the other method has an advantage on the delayed recall tests. The results also show that certain homonyms give better results when taught by one method and certain others when taught by the other method.

### The "Separate" and "Together" Methods Compared on Spelling Accuracy

The data in Tables III and IV are based on the number of right spelli[illegible] found on the

[18] Two-right spellings are those in [illegible] pupil spelled the homonyms right on both chances.

p. 269. Chicago: J. B. Lippincott and Company, 1924.

[20] Harold A. Edgerton and Donald G. Paterson, "Table of Standard Errors and Probable Errors of Percentages for Varying Numbers of Cases." *Journal of Applied Psychology*, X (1926), 378–391.

[21] Yule, *op. cit.*, p. 257.

immediate recall tests. Each pupil in the third grade studied thirty homonyms by the "separate" method and thirty by the "together" method. In the sixth grade each pupil studied forty-three homonyms by each method. The tables show results based on thirty homonyms in the third grade and on forty-three in the sixth grade, which are the number each pupil studied by each method. Each of the homonyms was tested twice in sentences, a procedure which gave the third grade pupils sixty chances to spell their thirty homonyms, and the sixth grade pupils eighty-six chances to spell their forty-three.

The main results of the study are based on the sum of the one-right and two-right spellings. The two-right spellings are those in which the pupil spelled the homonyms right on both chances. The one-right spellings are those in which the pupil was right only once.

A graphical representation of the results is given in Figure 1 which shows the average number of right spellings, errors, and confusions found in the third and sixth grades on the immediate and delayed recall tests.

Table III should be read as follows: In the third grade on the immediate recall test the 588 pupils studying by the "separate" method average 44.3214 right spellings out of sixty spelling opportunities; the 588 pupils studying by the "together" method average 45.8384 right spellings out of sixty spelling opportunities, etc.

TABLE III

THE SPELLING ACCURACY RESULTS ON THE IMMEDIATE RECALL TESTS

| | Third Grade | | Sixth Grade | |
|---|---|---|---|---|
| | Separate | Together | Separate | Together |
| No. of pupils | 588 | 588 | 636 | 636 |
| No. of homonyms in test | 30 | 30 | 43 | 43 |
| No. of spelling opportunities | 60 | 60 | 86 | 86 |
| A. M. right spellings[a] | 44.3214 | 45.8384 | 68.4403 | 73.0740 |
| S. D.[b] | 12.1613 | 12.2809 | 12.3464 | 10.9656 |
| "r"[c] | .6644 | .6497 | .7011 | .6378 |
| Diff. in A. M. | | 1.517 | | 4.634 |
| S. E. diff. (m)[d] | | .537 | | .484 |
| Critical ratio | | 2.82 | | 9.57 |

[a] Arithmetic mean.
[b] Standard deviation.
[c] "r" is the correlation between the pretest and the immediate recall test.
[d] Standard error of the difference for matched groups.

The results of Table III, based on the immediate recall tests, show a statistically significant difference (critical ratio 9.57) in the sixth grade, and a difference approaching significance (critical ratio 2.82) in the third grade for the "together" method. These results indicate that for immediate recall the "together" method is much better than the "separate" method.

The results for the pupils who took the delayed recall tests are given in Table IV.

TABLE IV

THE SPELLING ACCURACY RESULTS ON THE DELAYED RECALL TESTS

| | Separate | Together | Separate | Together |
|---|---|---|---|---|
| | Third Grade | | Sixth Grade | |
| No. of pupils | 297 | 295 | 279 | 279 |
| No. of homonyms in test | 30 | 30 | 43 | 43 |
| No. of spelling opportunities | 60 | 60 | 86 | 86 |
| A. M. right spellings[a] | 41.2088 | 41.0407 | 66.7419 | 64.8853 |
| S. D.[b] | 13.9465 | 13.7147 | 12.7817 | 13.7296 |
| "r"[c] | .7449 | .7381 | .7195 | .7546 |
| Diff. in A. M. | .168 | | 1.857 | |
| S. E. diff. (m)[d] | .763 | | .799 | |
| Critical ratio | .22 | | 2.32 | |

[a] Arithmetic mean.
[b] Standard deviation.
[c] "r" is the correlation between the pretest and the delayed recall test.
[d] Standard error of the difference for matched groups.

Fig. 1
Average Number of Right Spellings, Errors, and Confusions on the Immediate Recall and Delayed Recall Tests

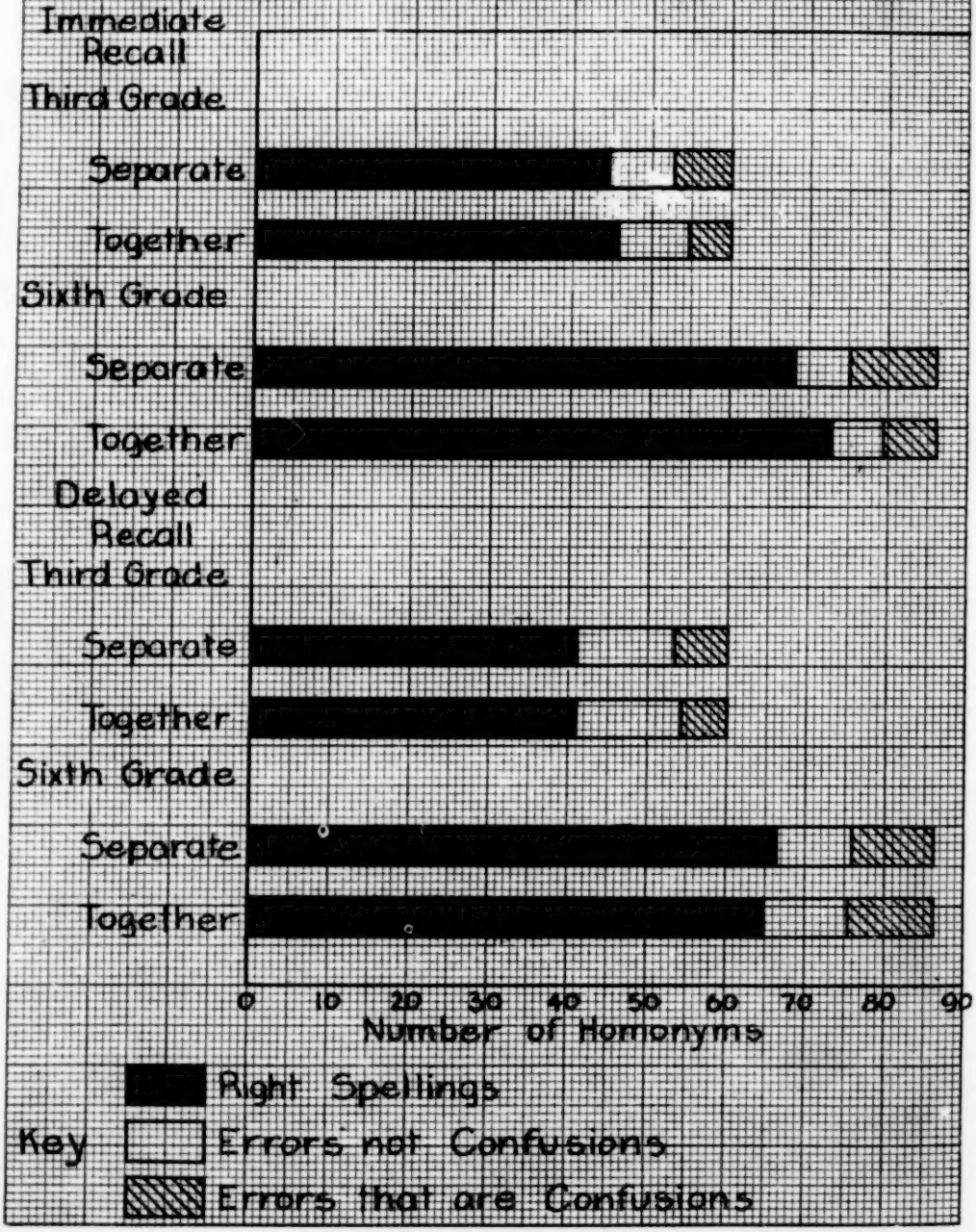


Table IV, which reveals the results on the delayed recall tests, shows a reverse trend from that in the preceding table, and in both grades gives an advantage to the "separate" method. A consistent, but not statistically significant, difference is found for the "separate" method in the third grade (critical ratio .22), and in the sixth grade (critical ratio 2.32) on the delayed recall test.

Since the final measure of spelling accuracy is permanent recall, greater weight is allowed the evidence on delayed recall than that on immediate recall. Considering the consistency of the findings on the delayed recall tests, a reasonable interpretation is to favor the "separate" method as giving slightly better results than the "together" method.

TABLE V

THIRD GRADE HOMONYM CLUSTERS GROUPED ACCORDING TO THE METHOD GIVING THE HIGHER SPELLING ACCURACY ON THE DELAYED RECALL TEST

| | Per cent Advantage | Critical Ratio | | Per cent Advantage | Critical Ratio |
|---|---|---|---|---|---|
| **Group A—Separate Method** | | | | | |
| close | 22 | 4.23 | meat | 0 | |
| clothes | 13 | 2.21 | meet | 2 | |
| | | | mete | 12 | |
| grate | 15 | 2.49 | | | |
| great | 9 | | pair | 12 | |
| | | | pare | 6 | |
| heard | 13 | 2.16 | pear | 10 | |
| herd | 21 | 3.49 | | | |
| | | | right | 9 | |
| knew | 18 | 3.03 | write | 24 | 4.03 |
| new | 20 | 3.65 | | | |
| | | | wares | 21 | 3.53 |
| knights | 24 | 3.99 | wears | 7 | |
| nights | 18 | 2.99 | | | |
| knows | 24 | 4.04 | | | |
| nose | 10 | | | | |
| **Group A—Together Method** | | | | | |
| aisle | 10 | | hours | 12 | |
| I'll | 12 | | ours | 19 | 3.77 |
| isle | 6 | | | | |
| | | | to | 17 | 3.16 |
| ant | 11 | | too | 26 | 4.97 |
| aunt | 8 | | two | 9 | |
| hears | 12 | | | | |
| here's | 14 | 2.60 | | | |
| **Group B—Separate Method** | | | | | |
| ate | 5 | | wait | 9 | |
| eight | 4 | | weight | 0 | |
| road | 6 | | weak | 4 | |
| rode | 1 (together) | | week | 8 | |
| their | 9 | | | | |
| there | 0 | | | | |
| **Group B—Together Method** | | | | | |
| billed | 8 | | hire | 8 | |
| build | 2 (separate) | | higher | 1 | |
| cents | 7 (separate) | | lets | 4 | |
| sense | 14 | 2.53 | let's | 7 | |
| fore | 9 | | some | 10 | |
| four | 10 | | sum | 5 | |
| **Group C** | | | | | |
| its | 6 (separate) | | mail | 3 (separate) | |
| it's | 6 (together) | | male | 7 (together) | |

### Confusion

The confusions which occur in spelling homonyms are a source of many errors. On the immediate recall test, the "separate" method has enough more confusions than the "together" method that the difference between them is statistically significant. On the delayed recall tests, the differences between the methods, although not statistically significant, show that the "separate" method has more confusions in the third grade and the "together" method has more in the sixth grade.

### The Method Giving Better Results for Each Homonym

The method which shows the better results for each homonym was determined as follows: The number of pupils who spelled each homonym correctly on the delayed recall test, on both of the opportunities to spell it, was determined for each method. This number was then converted into a percentage by dividing the number right by the number of pupils spelling the homonym by that method. The difference in percentage between the methods was then found for each homonym and com-

Table VI

Sixth Grade Homonym Clusters Grouped According to the Method Giving the Higher Spelling Accuracy on the Delayed Recall Tests

| | Per cent Advantage | Critical Ratio | | Per cent Advantage | Critical Ratio |
|---|---|---|---|---|---|
| **Group A—Separate Method** | | | | | |
| accepted | 7 | | companies | 15 | 2.56 |
| excepted | 21 | 3.67 | company's | 12 | 2.04 |
| adieu | 13 | 2.16 | fate | 5 | |
| ado | 5 | | fete | 11 | 1.87 |
| canvas | 11 | 2.12 | stationary | 13 | 2.24 |
| canvass | 5 | | stationery | 26 | 4.59 |
| choir | 12 | 2.17 | | | |
| quire | 1 | | | | |
| **Group A—Together Method** | | | | | |
| affect | 8 | | their | 3 | |
| effect | 13 | 2.50 | there | 6 | |
| | | | they're | 11 | 2.14 |
| council | 14 | 2.33 | | | |
| counsel | 6 | | | | |
| consul | 7 (separate) | | | | |
| **Group B—Separate Method** | | | | | |
| addition | 4 | | praise | 1 (together) | |
| edition | 5 | | prays | 6 | |
| | | | preys | 2 | |
| assistance | 5 | | | | |
| assistants | 0 | | profits | 3 | |
| | | | prophets | 6 | |
| attendance | 1 (together) | | | | |
| attendants | 7 | | rooted | 3 | |
| | | | routed | 4 | |
| billed | 1 | | | | |
| build | 5 | | sane | 7 | |
| | | | seine | 5 | |
| coarse | 8 | | | | |
| course | 5 | | soles | 0 | |
| | | | souls | 6 | |
| currants | 6 | | | | |
| currents | 3 | | vary | 2 | |
| | | | very | 5 | |
| ladies | 8 | | | | |
| ladies' | 7 | | waits | 6 | |
| lady's | 4 | | weights | 1 (together) | |
| | | | | | |
| passed | 9 | | weave | 6 | |
| past | 5 | | we've | 5 | |
| | | | | | |
| plain | 5 | | | | |
| plane | 9 | | | | |

Group B—Together Method

| | Per cent Advantage | Critical Ratio |
|---|---|---|
| baring | 13 | 2.16 |
| bearing | 6 (separate) | |

Group C

| | Per cent advantage Separate | Together |
|---|---|---|
| aisle | | 3 |
| I'll | 6 | |
| isle | | 4 |
| commence | | 1 |
| comments | 0 | |
| decent | | 1 |
| descent | 0 | |
| false | 0 | |
| faults | | 4 |
| fathers | 1 | |
| father's | | 1 |
| intense | | 7 |
| intents | 8 | |
| pause | 3 | |
| paws | 1 | |
| presence | | 3 |
| presents | | 1 |
| principal | | 4 |
| principle | 2 | |
| to | 3 | |
| too | | 1 |
| two | 1 | |
| wares | 4 | |
| wears | | 2 |
| writers | | 1 |
| writer's | 3 | |

pared, using the formula by Yule[22] to see if the difference between them is statistically significant. Critical ratios were determined for all the differences of thirteen per cent or more.

The homonyms taught in each grade are grouped into three divisions on the strength of the evidence for the "separate" or "together" method. Group A includes those homonym clusters in which one member of a cluster has eleven to twenty-six per cent advantage for one method or the other; group B includes those homonym clusters for which there is five to nineteen per cent advantage for the cluster for one method or the other; and group C includes those homonym clusters for which there is less than five per cent advantage for one method or the other.

The results of the delayed recall tests given in Table V show that 53.56 per cent of the third grade homonyms have appreciably higher accuracies, five to twenty-six per cent, when taught by the "separate" method and 39.28 per cent when taught by the "together" method.

The results of the delayed recall tests given in Table VI show that sixty per cent of the sixth grade homonyms have appreciably higher accuracies, five to twenty-six per cent, when taught by the "separate" method, and ten per cent when taught by the "together" method.

Gains in spelling accuracy during the delayed recall interval are recorded for nineteen homonyms in the third grade and thirty in the sixth grade. Of these homonyms more than three times as many were taught by the "separate" method as by the "together".

[22] Yule, *op. cit.*, p. 269.

### The Methods Compared for the Best and Poorest Pupils

The effectiveness of each method was determined for the highest and lowest ten per cent of the pupils. On the immediate recall tests the highest and lowest ten per cent of the pupils in both grades have a consistent, although statistically insignificant, advantage for the "together" method.

On the delayed recall tests the highest ten per cent of the pupils in the third grade do equally well by both methods and the lowest ten per cent of the pupils do slightly better by the "together" method. In the sixth grade on the delayed recall tests the highest and lowest ten per cent of the pupils have a small advantage for the "separate" method.

These findings are in agreement with the main findings of the investigation given in Tables III and IV. The only exceptions are found in the delayed recall tests in third grade which show that the highest ten per cent of the pupils do equally well by both methods and the lowest ten per cent do slightly better by the "together" method.

### Causes of Spelling Errors

Confusions account for fifty per cent of the errors found in approximately one third of

the homonyms taught in the third grade and in one half of those taught in the sixth grade. Another homonym of the cluster is the most frequent form of error found for approximately three fourths of the homonyms taught in third grade and in four fifths of those taught in sixth grade.

A careful examination of the tabulation of the misspelled forms shows that many of the errors were phonetic, that carelessness in writing, inversion, omission, and insertion of letters, misuse of the apostrophe, and omission of capital letters, as in other words, are common causes of the errors in spelling homonyms.

### Increases and Decreases in Spelling Accuracy and Confusions During the Delayed Recall Interval

One of the problems arising in this study was to account for the differences in loss in maintenance of spelling accuracy shown by the two methods for the period between the immediate and delayed recall tests. The evidence in both the third and sixth grades shows that the "together" method gives superior spelling results on the immediate recall tests. The delayed recall tests, given eight weeks later, show that the "separate" method has an advantage. Just what has happened to cause this reversal to take place? In Table VII are shown the decreases in spelling accuracy between the immediate and the delayed recall tests for the third and the sixth grades.

Table VII shows that the pupils taught by the "separate" method have 2.1234 homonyms per pupil smaller decrease in spelling accuracy in the third grade, and 3.7814 homonyms per pupil smaller decrease in the sixth grade than do those taught by the "together" method.

Table VIII presents the relation that was found to exist between the increases and decreases in spelling accuracy and confusions. Were the increases in errors which developed during the delayed recall interval caused by confusions?

A comparison of the increases in errors during the delayed recall interval given in Table VIII shows that the "together" method has 2.1234 more errors per pupil in the third grade and 3.7814 more in the sixth grade than does the "separate" method. Are these increases in errors caused by added confusions? If so, how do the two methods compare?

In the third grade, as shown by Table VIII the "separate" method has a decrease of .76 confusions per pupil (23.82 per cent of the increase in errors) during the delayed recall interval, and the "together" method has an increase of .43 per pupil (8.08 per cent of the increase in errors). The net result of this decrease in confusions in the "separate" method and the increase in confusions in the "together" method gives the "separate" method an advantage of having 31.9 per cent smaller increase in errors caused by confusions during the delayed recall interval than does the "together" method.

In the sixth grade, confusions cause 41.58 per cent of the increase in errors by the "separate" method and 55.10 per cent of the increase by the "together" method during the delayed recall interval. These results make

Table VII

The Increases and Decreases in Spelling Accuracy During the Delayed Recall Interval

| | Immediate Recall | | Delayed Recall | |
|---|---|---|---|---|
| | Separate | Together | Separate | Together |
| **Third Grade** | | | | |
| No. of pupils | 297 | 295 | 297 | 295 |
| No. of homonyms in test | 30 | 30 | 30 | 30 |
| No. of spelling opportunities | 60 | 60 | 60 | 60 |
| A. M. right spellings | 44.4040 | 46.3593 | 41.2088 | 41.0407 |
| Decrease between tests | | | 3.1952 | 5.3186 |
| Difference in decrease | | | 2.1234 | |
| **Sixth Grade** | | | | |
| No. of pupils | 279 | 279 | 279 | 279 |
| No. of homonyms in test | 43 | 43 | 43 | 43 |
| No. of spelling opportunities | 86 | 86 | 86 | 86 |
| A. M. right spellings | 68.7634 | 70.6882 | 66.7419 | 64.8853 |
| Decrease between tests | | | 2.0215 | 5.8029 |
| Difference in decrease | | | 3.7814 | |

TABLE VIII

THE RELATION BETWEEN ERRORS AND CONFUSIONS ON THE IMMEDIATE AND DELAYED RECALL TESTS

| | Immediate Recall | | Delayed Recall | |
|---|---|---|---|---|
| | Separate | Together | Separate | Together |
| **Third Grade** | | | | |
| No. of pupils | 297 | 295 | 297 | 295 |
| No. of homonyms in tests | 30 | 30 | 30 | 30 |
| No. of spelling opportunities | 60 | 60 | 60 | 60 |
| A. M. Errors | 15.596 | 13.6407 | 18.7912 | 18.9593 |
| A. M. Confusions | 7.3502 | 5.4644 | 6.5926 | 5.8881 |
| Confusions divided by errors[a] | 47.11 | 40.02 | 35.07 | 31.65 |
| A. M. increase in errors | | | 3.1952 | 5.3186 |
| Diff. in increase | | | | 2.1234 |
| A. M. increase in confusions | | (decrease) | .76 | .43 |
| Increase in confusions divided by increase in errors | | (decrease) | 23.82 | 8.08 |
| Per cent advantage | | | 31.90 | |
| | Immediate Recall | | Delayed Recall | |
| | Separate | Together | Separate | Together |
| **Sixth Grade** | | | | |
| No. of pupils | 279 | 279 | 279 | 279 |
| No. of homonyms in tests | 43 | 43 | 43 | 43 |
| No. of spelling opportunities | 86 | 86 | 86 | 86 |
| A. M. Errors | 17.2366 | 15.3118 | 19.2581 | 21.1147 |
| A. M. Confusions | 9.4265 | 7.3548 | 10.2688 | 10.5699 |
| Confusions divided by errors[a] | 54.69 | 48.00 | 53.32 | 50.07 |
| A. M. increase in errors | | | 2.0215 | 5.8029 |
| Diff. in increase | | | | 3.7814 |
| A. M. increase in confusions | | | .84 | 3.22 |
| Increase in confusions divided by increase in errors | | | 41.58 | 55.10 |
| Per cent advantage | | | 13.52 | |

[a] Per cent of the errors which are confusions.
[b] Per cent of the increase in errors which are caused by increase in confusions.

it clear that confusions in the "together" method during the delayed recall interval cause 13.52 per cent more of the increase in errors than they do for the "separate" method.

It may reasonably be assumed, then, that the superiority in maintenance for the "separate" method is caused by the smaller increase in errors which occurred during the delayed recall interval. This smaller increase in errors is partly accounted for in the third grade by the disappearance of confusions already present, and in the sixth grade by fewer confusions developing during the delayed recall interval.

## DISCUSSION OF FINDINGS

The evidence points clearly to two seemingly contradictory findings:

1. The "together" method is superior on immediate recall.

2. The "separate" method has an advantage on delayed recall.

The results offer an explanation of the findings of Pearson,[23] who discovered in 1911 that the "together" method was better, and Finkenbinder,[24] who found in 1923 that the "separate" method was better. The published material on Pearson's study shows that he gave his final test ten days after the instruction period ended. This would hardly measure all the loss that might occur if a longer delayed recall interval were employed.

Finkenbinder gave an immediate recall test at the end of the instruction period and a delayed recall test one month later.

The results of the studies by Pearson and Finkenbinder and those of this investigation lend themselves to consistent inference. They indicate: (1) that there is something inherent in the "together" method which causes the pupils to spell better on immediate recall than pupils who studied by the "separate" method; and (2) that there is

[23] Pearson, *op. cit.*
[24] Finkenbinder, *op. cit.*

something inherent in the "separate" method which enables pupils to retain their spelling ability better during a delayed recall interval than do the pupils who study by the "together" method.

The results show that the "together" method has a greater increase in confusions and errors than does the "separate" method during the delayed recall interval (Table VII). This fact indicates that the passage of time adversely affects the pattern of a homonym cluster when taught by the "together" method more than it does when taught by the "separate" method. It appears that the learning of two or more homonyms at the same time (together method) sets up patterns in the nervous system of which an important phase is the contrasts. These contrasts are made up of word pictures, meaning, spelling, use in sentences, etc., and are tied together by sound and time sequence. Such contrasts appear to facilitate immediate learning as is illustrated by such words as *fate-fete, canvas-canvass, their-there*. The distinctions are relatively easy to remember. With a lapse of time, however, these contrasts fade, forgetting takes place, and the differences in meaning or use become shadowy or even transpose themselves, resulting in an increase in confusions as well as in errors.

On the other hand, the "separate" method sets up a pattern in the nervous system centered around one word with its sound, meaning, use, spelling, etc., all forming radiations of the pattern. When another member of the cluster is studied it may or may not be tied to the first in the pattern. Frequently it is difficult to bring the members of a cluster together in meaning, as use so dominates the patterns of sound that the right form appears upon stimulus with no confusions resulting. The distinctions between members of a cluster are not stressed as such. Since the patterns are more isolated, the sound may call up only one response, thus causing the pupil to use the same spelling whenever that sound appears. This explains how confusions may occur, and it may also account for the finding that the "separate" method has more confusions on the immediate and delayed recall tests than has the "together" method.

In the "separate" method the lapse of time affects all the patterns of the homonym cluster, the patterns of each homonym fading rather than the differences between them fading and causing them to be confused. This offers a reasonable explanation of why the "separate" method has fewer new confusions than the "together" method on the delayed recall tests.

In the "separate" method the pattern has opportunity for additional enrichment. In comparison, the pattern for a cluster of homonyms set up by the "together" method is centered around the cluster and is largely based on the contrasts between the homonyms. Opportunities for strengthening these contrasts by incidental learning are few as the homonyms appear in sentences separately. The net effect, then, is for the incidental learning of these homonyms in their reading, language, writing, and other subjects to strengthen the "separate" method more than the "together".

This appears to be a reasonable explanation of the fact that the "separate" method had over three times as many homonyms showing gains in spelling accuracy during the delayed recall interval as did the "together" method.

The "separate" method, then, appears to be a better method for the teaching of homonyms for permanent recall because it depends upon an enriched individual pattern for each word rather than upon contrasts between the members of a cluster.

Many of the homonyms appear in the child's vocabulary before he starts to school. The best method for the teaching of these homonyms will probably have to be determined for each homonym. This study indicates the method which gives better results for a number of them.

The spelling of those homonyms appearing in the child's vocabulary after he starts to school is probably best taught when they appear. This method would enable some of the homonyms to appear in lessons several grades apart, which would seem desirable as it would enable the pupil to learn one of the homonyms, and after its spelling and use are well fixed, to learn the other one.

The slightly better results found in this study for the "separate" method came as a result of teaching the homonyms in lessons with approximately one week intervening between the study of one homonym and the appearance of the other member of a cluster.

It is the opinion of the author that if the experiment could have been set up so that a longer time, for example, one semester, had intervened between the study of the homonyms, the results would have shown that the "separate" method had fewer confusions and larger differences in spelling accuracy in its favor.

The possibility should not be overlooked that a combination of the two methods, "separate" and "together", might give better results than either alone. At least, it appears that if the homonyms are introduced separately and confusions appear, then they may be brought together with profit, to contrast their differences and to set the pupil right. This phase of the problem of homonym instruction might profitably be investigated further.

On the other hand, if the spelling of these homonyms is to become a fixed habit, so that when one needs to use them no doubt will arise as to which homonym to spell, the best way psychologically seems to be to fix them through use in the setting in which they will be called for. This would suggest the "separate" method.

## Conclusions

Based on the evidence for the specific homonyms included in this study the following conclusions are believed to be warranted.

1. The "together" method results in greater immediate spelling accuracy than does the "separate" method. In the third grade the difference was not statistically significant (critical ratio 2.82) but in the sixth grade the difference was statistically significant (critical ratio 9.57). This probably means that the pupils who studied these homonyms by the "together" method, which encouraged contrasts between them, were able to acquire the right spelling and hold it for a short time better than could the pupils taught by the "separate" method, which encouraged them to learn each word by itself.

2. A consistent advantage, although not statistically significant, was found in the third grade (critical ratio .22) and in the sixth grade (critical ratio 2.32) for the "separate" method on the delayed recall tests. The author believes that the final measure of spelling accuracy is permanent recall, and for that reason gives greater weight to the evidence on delayed recall than to that on immediate recall. Without statistically significant evidence, a final decision as to the method better for delayed recall should be withheld. Although not statistically significant, the weight of evidence now available on the teaching of homonyms for permanent recall indicates that the "separate" method on the average is better than the "together" method. Exceptions must be made in the cases of certain homonyms since they give better results when taught by the "together" method.

The pupils who learn the homonyms by the "together" method hold them for a time by the contrasts between them; after a period of time elapses, these contrasts become indistinct and hazy and forgetting occurs. The pupils who learn the homonyms by the "separate" method hold them by patterns which are centered around the homonym. The patterns fade as a whole, thus causing forgetting, but not so much as results from the "together" method.

3. The "separate" method has enough more confusions than the "together" method on the immediate recall tests in the third and sixth grades that the difference between them is statistically significant. On the delayed recall test, the difference in the number of confusions, although not significant, shows that the "separate" method has more confusions in the third grade and the "together" method has more in the sixth grade.

4. A small and consistent difference, although statistically insignificant, was found in both grades for the highest and lowest ten per cent of the pupils for the "together" method on the immediate recall test. On the delayed recall test, the highest ten per cent of the third grade pupils did equally well by both methods, and the lowest ten per cent of the pupils did slightly better by the "together" method. In the sixth grade the highest and lowest ten per cent of the pupils had a small advantage, not statistically significant, for the "separate" method.

5. Confusions account for fifty per cent of the errors found in approximately one third of the homonyms taught in the third grade and in one half of those taught in the sixth grade. Another homonym of the cluster is the most frequent form of error found for approximately three fourths of the homonyms taught in third grade and in four fifths of those taught in sixth grade.

6. The superiority in maintenance of spelling accuracy found for the "separate" method is accounted for by the smaller increase in errors which occurred during the delayed recall interval. This smaller increase in errors is partly accounted for in the third grade by the disappearance of confusions already present, and in the sixth grade by fewer confusions developing during the delayed recall interval.

7. Gains in spelling accuracy during the delayed recall interval are recorded for nineteen homonyms in third grade and thirty in sixth grade. Of these homonyms more than three times as many were taught by the "separate" method as by the "together".

8. Of the sixty homonyms taught in the third grade, an eleven to twenty-six per cent advantage for one member of a cluster was found for teaching twenty-two of them by the "separate" method and twelve by the "together" method. Of the group which showed from five to ten per cent advantage, ten gave better results by the "separate" and twelve by the "together" method. The remaining four showed little advantage for either method.

The third grade homonyms which showed strongest evidence for the "separate" method are:

| | | | |
|---|---|---|---|
| close | knew | meat | right |
| clothes | new | meet | write |
| | | mete | |
| grate | knights | | wares |
| great | nights | pair | wears |
| | | pare | |
| heard | knows | pear | |
| herd | nose | | |

The third grade homonyms which showed strongest evidence for the "together" method are:

| | | |
|---|---|---|
| aisle | hours | to |
| I'll | ours | too |
| isle | | two |
| | ant | |
| hears | aunt | |
| here's | | |

Of the eighty-six homonyms in the sixth grade, an eleven to twenty-six per cent advantage for one member of the cluster was found for teaching fourteen of them by the "separate" method and seven by the "together" method. Of the group which showed from five to ten per cent advantage for the cluster, thirty-seven of them showed better results by the "separate" method and two by the "together" method. The remaining twenty-six showed little advantage for either method.

The sixth grade homonyms which showed strongest evidence for the "separate" method are:

| | | | |
|---|---|---|---|
| accepted | canvas | companies | stationary |
| excepted | canvass | company's | stationery |
| adieu | choir | fate | |
| ado | quire | fete | |

The sixth grade homonyms which showed strongest evidence for the "together" method are:

| | | |
|---|---|---|
| affect | council | their |
| effect | counsel | there |
| | | they're |

The author has been unable to develop from the data available any satisfactory hypothesis which would explain the fact that certain homonyms show higher spelling accuracies when taught by the "separate" method, while others have higher accuracies when taught by the "together" method. There is perhaps a common element or principle which, if known, would explain the behavior of these homonyms. Two possible approaches for its discovery would be: (1) to repeat such a study as this with a different set of homonyms and by comparison with the results of this study attempt to find a common element between them; (2) to follow by means of case studies the experiences of the pupils in learning to spell the homonyms by each method.

Finally, the observation should be made that the pupils learned to spell by both methods, and that such average differences as existed between the methods were small.

## Bibliography

Ashbaugh, E. J. *The Iowa Spelling Scales.* Journal of Educational Research Monographs, No. 3. Bloomington, Illinois: Public School Publishing Co., 1922.

Burnham, Wm. H. "The Hygiene and Psychology of Spelling," *Pedagogical Seminary,* XIII (December, 1906), 495.

Cooper, E. J. "The 1,035 Words Most Commonly Misspelled by Minnesota Pupils in Grades Three to Eight." Master's Thesis, University of Minnesota, 1927.

Edgerton, Harold A., and Paterson, Donald G. "Table of Standard Errors and Probable Errors of Percentages for Varying Numbers of Cases," *Journal of Applied Psychology,* X (1926), 378–391.

Fallows, Samuel. *A Complete Dictionary of Synonyms and Antonyms.* Chicago: F. H. Revell Company, 1883.

Finkenbinder, E. O. "The Spelling of Homonyms: An Experimental Investigation of Teaching Them," *Pedagogical Seminary,* XXX (September, 1923), 241–251.

Fitzgerald, James A. "Words Misspelled Most Frequently by Children of the Fourth, Fifth, and Sixth Grade Levels in Life Outside the School," *Journal of Educational Research,* XXVI (November, 1932), 213–218.

Foran, Thomas George. *The Psychology and Teaching of Spelling,* pp. 40–41. Washington, D. C.: The Catholic Education Press, 1934.

Horn, Ernest, and Ashbaugh, E. J. *The Horn–Ashbaugh Fundamentals of Spelling.* Chicago: J. B. Lippincott and Company, 1928.

Lindquist, E. F. "The Significance of A Difference Between Matched Groups," *Journal of Educational Psychology,* XXII (March, 1931), 197–204.

Masters, H. V. *A Study of Spelling Errors.* University of Iowa Studies in Education, IV, No. 4. Iowa City, Iowa, 1927.

Pearson, Henry C. "The Scientific Study of the Teaching of Spelling," *Journal of Educational Psychology,* II (May, 1911), 241–252.

Peterson, Evelyn F. "An Analysis of Spelling Errors." Master's Thesis, University of Iowa, 1930.

Webster, Noah. *Webster's New International Dictionary of the English Language,* p. 1194. (2nd ed., unabridged). Springfield, Mass.: G. and C. Merriam Company, 1934.

Yule, G. U. *An Introduction to the Theory of Statistics,* pp. 257, 269. Chicago: J. B. Lippincott and Company, 1924.

# A COMPARISON OF FOUR TYPES OF SPELLING TESTS FOR DIAGNOSTIC PURPOSES

JOE E. MOORE
*George Peabody College for Teachers*

The spelling ability of children is determined ordinarily by tests of some kind. The basic assumption behind most of these diagnostic tests is that they will determine whether or not a child can spell a word in a specific situation. In turn, the way a spelling test predicts a pupil's ability to spell a word depends upon two factors: the reliability of the test as a diagnostic instrument and the variability of the subject taking the test. Obviously, then, it is important to know what type of test should be applied in order to obtain the most reliable measure of a pupil's misspelling.

The purpose of this study was to investigate the variability of sixth grade pupils on four kinds of spelling tests. The test words and the procedure used in this study were taken from a previous study by Northby[1] who investigated five types of spelling tests. The oral form used by Northby was omitted in the present study.

The subjects consisted of 102 high-sixth grade pupils from three public schools in Nashville, Tennessee: 38 pupils from the Tarbox School, 40 from Jones School, and 24 from Clemons School. The entire high-sixth grade enrollment, including both boys and girls, in each of the schools investigated was used in this study.

The twenty words used in the list presented in this study were selected by Northby from the Iowa Spelling Scales developed and published by E. J. Ashbaugh.[2] With respect to these words Northby says, "Approximately half of the twenty words were of fifty per cent difficulty for average sixth graders, one-fourth of greater than fifty per cent difficulty, and one-fourth less than fifty per cent difficulty."[3] The list of words used by Northby, together with the per cent of average sixth grade pupils spelling them correctly, is given below:

| | |
|---|---|
| ninety (54%) | orchestra (55%) |
| succeeded (54%) | occasionally (33%) |
| similar (54%) | convenience (23%) |
| nickel (51%) | exhibition (23%) |
| admission (50%) | schedule (9%) |
| carnival (50%) | arriving (63%) |
| appearance (51%) | various (59%) |
| excitement (50%) | humor (78%) |
| planned (45%) | Saturday (87%) |
| finally (44%) | island (93%) |

Since the tests employed in this analysis were the same that Northby used in his study, it is obvious that the present investigation follows the earlier study in purpose. Both seek to determine whether there is any basis for anticipating a significant difference in favor of any one of these four most widely used forms of presenting spelling materials. Probably the notable difference in the two investigations is that, whereas 102 children were included in the present study, Northby made his comparisons with only 43 cases.

A brief description of the four tests, as developed by Northby, is as follows:

1. Story form: As the title suggests, the twenty words were worked into a short story which follows:

All week we had *planned* to go to the *carnival* on *Saturday*. It seemed that the day would never come, but *finally* it did. Right on *schedule* that morning about *ninety* of us *succeeded* in crowding into the cars that had been provided for our *convenience* to take us to the *island*.

Everyone was in a good *humor*. Upon *arriving* there we found *various* shows and many fine *exhibitions* to attract attention. For some of these the *admission* price was a *nickel*, but *occasionally* we found some that were free. The *appearance* of several clowns added greatly to the *excitement*. No two were *similar*.

There was much music, but we liked one *orchestra* better than the others.

2. Time dictation tests: Though the time dictation tests used in this investigation were

[1] Arwood S. Northby, "A Comparison of Five Types of Spelling Tests for Diagnostic Purposes," *Journal of Educational Research*, XXIX (January, 1936), 339–46.

[2] E. J. Ashbaugh, *Iowa Spelling Scales*. Bloomington, Illinois: Public School Publishing Co., 1922.

[3] *Op. cit.*, p. 341.

also the same as those used by Northby, the method of presenting them differed slightly due to the conditions under which the present study was made. The tests used were given to the subjects in a large study hall where 75 or more other pupils were preparing their lessons and not taking the spelling test. As a result it was difficult to make the sentences understood without repeating them. Each sentence in the dictation was pronounced twice rather than once (as in Northby's study), and the time limits given herewith were checked on a stop watch. As soon as the signal *stop* was given the next sentence was immediately pronounced twice, and the time limit determined for it was observed. It will be seen that according to this method the pupils had a longer pause than would be the case in the continuous dictation observed in the formal time-dictation sentences, but the actual time used in writing, as determined by Brueckner and Melby[4] and adopted by Northby, is the same.

The twelve sentences used in this test, with the time limit for each sentence, are as follows:

| | Number of letters | Time in seconds |
|---|---|---|
| A *nickel* was the *admission* price | 27 | 25 |
| On *Saturday* we *planned* a trip | 24 | 23 |
| The *island* was *ninety* miles wide | 27 | 25 |
| The *exhibition* finally began | 25 | 23 |
| The *orchestra* played *various* kinds of music | 37 | 34 |
| We saw the *schedule* of *arriving* trains | 32 | 30 |
| The *carnival* was *similar* to many others | 33 | 31 |
| Bad *humor* was *occasionally* shown | 28 | 26 |
| John *succeeded* in winning the prize | 30 | 28 |
| His *appearance* caused much laughter | 31 | 29 |
| The *excitement* was great | 21 | 20 |
| It was *planned* for the *convenience* of the people | 40 | 37 |

3. List form: The list of words used in this section of the test was the same as that previously given. The method of giving the test was the same as that used by Northby: each word was pronounced twice before the pupils wrote.

4. Multiple choice: The multiple choice form was the same as that used by Northby. A mimeographed sheet containing the four most common misspellings of each word was included with the correct spelling. The pupil had five possible choices for each word. The subject responded by making a mark under the word that he felt was the correct spelling for the one represented in each line. The following material will enable the reader to get a clear idea of what the multiple choice section was like:

nickle, nickel, nicle, nickele, nichle
humer, humor, humur, humar, hummor

*Procedure*

The four tests were administered one after the other in the order here presented, on the same day, approximately thirty-five minutes being required to take the entire set of tests. The writer administered all tests in order to hold constant the factor of pronunciation.

*Results*

Table I shows that when the criterion of difficulty is applied, the story form is perhaps the best measure of a pupil's spelling ability. The difference in the story form over timed dictation is slight.

The differences between the means were checked to see if they were significant by dividing each difference by its Standard Error of the Difference. The differences between the means, and differences divided by the Standard Error of the Differences are presented in Table II.

TABLE I

MEAN SCORES AND STANDARD DEVIATIONS ON FOUR TYPES OF SPELLING TESTS

| Type of Test | Mean Score | Standard Deviation | Northby's Data |
|---|---|---|---|
| Story Form | 6.62 | 3.52 | 8.51 |
| Timed Dictation | 7.24 | 3.70 | 8.32 |
| List Form | 7.90 | 3.50 | 10.02 |
| Multiple Choice | 11.36 | 3.46 | 13.83 |

[4] Leo J. Brueckner and Ernest O. Melby, *Diagnostic and Remedial Teaching*, pp. 385–87. Boston: Houghton Mifflin Co., 1931.

TABLE II

DIFFERENCES BETWEEN THE MEANS OF THE FOUR SPELLING TESTS TOGETHER WITH THE STANDARD ERRORS OF THE DIFFERENCES BETWEEN MEANS, CORRELATED MEASURE, AND THE RATIOS BETWEEN THE TWO*

| Type of Test | $M_1 - M_2$ | S. D. of Diff. | Difference / S. D. of Difference |
|---|---|---|---|
| Multiple Choice ---- Story Form | 4.74 | .49 | 9.67 |
| Multiple Choice ---- List Form | 3.46 | .49 | 7.06 |
| Multiple Choice ---- Timed Dictation | 4.12 | .50 | 8.21 |
| Story Form ------- List Form | 1.28 | .50 | 2.58 |
| Story Form ------- Timed Dictation | .62 | .51 | 1.22 |
| List Form --------- Timed Dictation | .66 | .51 | 1.29 |

* In the calculation of the means and Standard Deviation a step interval of 2 was used.

Whenever a difference between two means is equal to or greater than three times its Standard Error of the Difference, this difference is considered to be statistically significant. In only three of the six instances is the difference between the several means equal to or greater than three times its Standard Error. In Northby's study, to the contrary, the differences obtained were reliable in all except one instance, that being the difference between the story form and the timed dictation form. Incidentally, it will be seen from Table II that in the present study the difference between the story form and the timed dictation form is less reliable than any of the others. There is another marked difference in the results obtained in this study and those obtained by Northby. Northby found a statistically reliable difference between the list form and the timed dictation. The differences obtained in this study between the story form and the list form, and between the list form and timed dictation were not reliable.

The percentage of the 102 pupils spelling each word correctly, tabulated according to the type of test, is given in Table III. The results obtained by Ashbaugh from Iowa pupils in the sixth grade are given at the right under the column labeled "Iowa Standards."

Since Ashbaugh's list was given by the list form method, it should be compared with that method as used in the present study. It will be seen from Table III that there is an appreciable difference between Ashbaugh's results and the results obtained in the present study when the list form was used. In only four words of the twenty did the present sample of Tennessee children exceed the Iowa group in correct spelling. It is impossible to account for the difference unequivocally. However, it may have resulted from a difference in textbooks, the amount of emphasis

TABLE III

PERCENTAGE OF PUPILS SPELLING CORRECTLY THE TWENTY WORDS IN THE FOUR TESTS

| Word | Story Form | Dictation | List | Multiple Choice | Iowa Standards* |
|---|---|---|---|---|---|
| ninety | 50.0 | 54.9 | 66.6 | 66.6 | 54 |
| succeeded | 17.6 | 19.6 | 21.5 | 34.3 | 54 |
| similar | 6.8 | 5.9 | 9.8 | 20.6 | 54 |
| nickel | 42.3 | 45.1 | 45.1 | 61.8 | 51 |
| admission | 52.9 | 57.9 | 54.9 | 42.3 | 50 |
| carnival | 35.3 | 36.3 | 37.2 | 49.0 | 50 |
| appearance | 9.8 | 6.8 | 11.7 | 11.7 | 51 |
| excitement | 30.4 | 33.3 | 36.3 | 64.7 | 50 |
| planned | 65.7 | 64.7 | 70.6 | 80.4 | 45 |
| finally | 50.9 | 63.7 | 60.8 | 71.6 | 44 |
| occasionally | 1.0 | 2.9 | 2.9 | 49.0 | 33 |
| convenience | 1.0 | 1.9 | 1.9 | 21.6 | 23 |
| orchestra | 3.9 | 4.9 | 5.9 | 42.3 | 55 |
| exhibition | 1.9 | 4.9 | 5.9 | 42.3 | 23 |
| schedule | 7.8 | 6.8 | 6.8 | 49.0 | 9 |
| arriving | 33.3 | 57.8 | 73.5 | 80.4 | 63 |
| various | 17.6 | 22.5 | 25.5 | 41.2 | 59 |
| humor | 28.4 | 29.4 | 36.3 | 48.0 | 78 |
| Saturday | 58.8 | 58.8 | 70.6 | 91.2 | 87 |
| island | 87.3 | 90.2 | 92.2 | 96.1 | 93 |

*Determined by list form method.

placed upon spelling, divergent methods of teaching spelling, difference in the time of the school year when the two experiments were conducted, variations of conditions under which the studies were made, or from combinations of these factors.

In the correct spelling of the twenty words the story form has the lowest percentage of correct spelling in fourteen instances. The multiple choice form has the highest percentage of correct spelling of the four tests given. In this connection it should be remembered that the multiple choice form was the only recognition test given. Of the three recall tests the list form shows the highest percentage of correct spelling.

The total number of different words spelled by all the subjects was 2040. Since these words were spelled in four different situations, the total number of spelling attempts by all pupils was 8160. Of these, 5990 words were misspelled. The number of misspellings, classified as to forms, are presented in Table IV according to type of test.

TABLE IV

NUMBER OF TIMES THE WORDS WERE MISSPELLED ON EACH OF THE FOUR FORMS

| Story Form | Timed Dictation | List Form | Multiple Choice | Total Errors |
|---|---|---|---|---|
| 1415 | 1359 | 1289 | 927 | 5990 |

The data indicate that among these four types of tests pupils made most errors on the story form of spelling test and fewest errors on the multiple choice form. However, since the trend in reduction of errors is consistent throughout the four tests in the order given to the subjects, there may be some indications that there was a practice effect operating throughout the series. Northby's results[5] show a similar trend in reduction of errors with the exception of the timed dictation form, which he found slightly higher in errors than the story form. It is generally well-known that the recognition type of test seems to be much easier than the recall type. This fact may explain, at least partially, the marked decrease of errors in this type of test as shown in Table IV.

*Conclusion*

There is a marked variability of pupils on the several tests. Some of the words found difficult on one form of test seem easier when presented in another form. The story and the timed dictation types of tests appear most difficult. The multiple choice form is easiest for the majority of students. From the results on the multiple choice form of the test it would seem that this test is either sampling an ability that none of the other tests is measuring, or else shows the effect of practice on the test results. From the previous findings in the field of educational research it would appear that perhaps the former possibility is the more valid assumption to apply here. Further research should be done with this type of spelling test.

If the difficulty of the test is to be used as a basis for its validity, then the story form and timed dictation form would be the best for diagnostic testing. The variability (Standard Deviation) around the mean of the story form and of the timed dictation form is about the same, being 3.54 in the former case and 3.70 in the latter. The Standard Deviation around the mean of the timed dictation form is greater than the variability on any of the other tests. The normality of the data in this study is evidenced by the unusual uniformity of both the means and the Standard Deviation of the means. When it is borne in mind that the pupils represented in this study came from three sections of the city of Nashville, Tennessee, representing a fair cross-section of the city, the normality seems even more significant.

Other than the multiple choice form the list form has the highest average and the smallest Standard Deviation for correct spellings. If a pupil could spell at all it would seem that his performance could be measured with greater reliability on the list form than on any of the others. Since the differences between the list form, the story form, and the timed dictation tests are not reliable, the list form could be used with assurance that it would be as accurate a measure of the possible errors as any of the other tests given. However, it should be pointed out that most of the pupils have been using the list form in spelling, which fact may have something to do with the consistency of the measures of central tendency and variability.

These data in general verify the conclusions of Northby who made use of the forms used in this study as well as an additional oral test. The superiority of Northby's group, as represented by mean scores, over the group

[5] *Op. cit.*, p. 344.

in this study would tend to indicate the presence of some selective factor or factors.

These data indicate that different spelling tests tend to test different types of ability. A pupil may be able to recognize the correct spelling on the multiple choice form but not be able to spell it on any of the three recall tests. Furthermore, the results of this study seem to show that the situation in which a pupil used a word will make an important difference in his ability to spell it correctly.

The findings in this investigation should be suggestive for further research. One question that arises in this study is to determine whether or not there is a statistically reliable difference between the timed dictation test used in this study and the one used by Northby. A similar test should be constructed which would parallel the particular course of study in each state to see if the data support the conclusions of this study and of Northby's investigation. The fact that pupils do so much better on the multiple choice type of test might cause a marked influence on the other forms, should it be given first. The order of presenting the different forms of the test might cause a marked difference in results.

# A STUDY OF STATUS AND IMPROVEMENT OF COLLEGE FRESHMEN IN CERTAIN SKILLS OF ENGLISH COMPOSITION†

WILMA LESLIE GARNETT
*Kent State University, Kent, Ohio*

A study of the status of proficiency in English attained by students attending institutions educating teachers seems quite imperative at this time in order that a beginning point may be found from which to undertake obviously needed improvements in the programs of English teaching offered by these institutions. It is a matter of rather general knowledge that the status reached in English compositional skills is one of low proficiency for the majority of college freshmen. If the teacher-training institutions of the country would make a concerted attack upon raising this status for the students whom they are to send out to be teachers in the elementary and secondary schools, much could be done to raise the level of proficiency in English skills throughout this country.

A somewhat expository statement of the problem of this entire study or investigation is this: a study of the status in basic skills in written English attained by freshmen in three teachers colleges, together with suggestions growing out of the implications of that study for certain revisions in the course in English composition on the college level.

A statement of the problem in terms of some of the specific objectives is given in order to define in a more specific way the delimitations accepted.

*1. To carry out a satisfactory testing program in order to discover the status of proficiency in certain basic skills in written English attained by all groups in the experiment

2. To study compositions, tests, and other material written by the students tested

*3. To make a detailed inventory of the results of the testing program and of the composition materials selected for that purpose

*4. To analyze the results of the tests and to analyze the compositions for types of errors in basic skills.

*5. To interpret the results of the program of testing and composition writing

6. To evaluate the results of the entire program

7. To develop from the inventory taken, techniques useful in making an inventory of the needs of college freshmen in basic skills in English

8. To carry on, at one of the colleges, during the school year 1931–1932, a remedial program with the group of about twenty students ranking lowest on the inventory tests of basic skills in written English

9. To develop from the work with the remedial group procedures which should be effective with any group of freshmen weak in basic skills in written English

10. To evaluate the remedial program

*11. To work out suggestions from the evaluations of the testing program and the remedial program of teaching for certain revisions in the course in English composition for teachers colleges

The present study is made in an attempt to help raise the standards of training in the basic skills in written English to a higher level of proficiency. The data collected in this study yield indications in favor of a program of self-instruction in English composition on the college level.

This study of status was carried on in three teachers colleges in two states of the Middle West. The three colleges will be designated in the thesis as College A, Col-

† An abstract of part of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Department of Education, Graduate College of the State University of Iowa, 1937.

The present study had its origin in a suggestion given by Dr. Ernest Horn. To him and to Dr. Harry A. Greene, the writer of this thesis desires to express sincere gratitude for their guidance in directing the study.

The writer is grateful to the several institutions which cooperated so generously in carrying on this investigation.

To others, the writer is indebted for generously giving of their time and services: Dr. Helen K. Mackintosh, Mrs. James McKenzie, Dr. Luella Wright, Dr. W. R. Lueck, Dr. M. F. Carpenter, and Dr. Walter W. Cook.

* The present abstract deals with phases of the objectives starred.

lege B, and College C, respectively. The group of sixteen freshmen from College C chosen for special study will be called the Remedial Group. As a point of departure for the study of status, a summary of certain points of general information about the three institutions concerned is presented in Table I.

TABLE I

NUMBER OF STUDENTS TESTED AT THE THREE COLLEGES

| College | Year | Group | Number of Cases in the Experiment |
|---|---|---|---|
| A (Four-year school) | 1931-1932 | Freshmen | 54 |
| | | Sophomores | 31 |
| | | Juniors | 29 |
| | | Seniors | 35 |
| B (Four-year school) | 1931-1932 | Freshmen | 62 |
| | | Sophomores | 14 |
| | | Juniors | 10 |
| | | Seniors | 8 |
| C (Two-year school | 1931-1932 | Freshmen (General Group) | 109 |
| | | Freshmen (Remedial Group) | 16 |
| | | Sophomores | 24 |
| C | 1934-1935 | Freshmen | 109 |
| | | Sophomores | 24 |
| | 1931-1932 Totals | Freshmen | 241 |
| | | Sophomores | 69 |
| | | Juniors | 39 |
| | | Seniors | 43 |
| | 1931-1932 | Total | 392 |
| | 1934-1935 | Total | 133 |
| | | Grand Total | 525 |

In this study, conducted along the lines of a combination of the experimental, comparative, and survey techniques, College C carried on the most extensive program. The other colleges carried on less extensive programs of testing, useful mainly as checks on the program conducted at College C.

The study was begun in the Fall of 1931. During the school year of 1931–1932, Fall and Spring testing programs in basic skills in written English were conducted at the three colleges named. In 1933, an extension of the study was planned for the school year 1934–1935 at College C. It was believed that by carrying on a similar testing program in 1934–1935, some comparisons might be drawn between the scores made at College C in that year and those made there in 1931–1932.

The tests chosen for the main experiment were these:

*Kuhlmann–Anderson Intelligence Tests, Grade IX—Maturity*

*Otis Self-Administering Tests of Mental Ability, Form A*

*Iowa Placement Examinations of English-Aptitude, Series 1, Revised, Forms A and B*

*Iowa Placement Examinations of English-Training, New Series, Forms X and Y*

*O'Rourke Grammar Survey Tests, Forms A and B, Tests 1, 2, and 3*

*Cook Spelling Test, Forms A and B*[1]

A phase of the program of 1931–1932 was projected into 1934–1935, at College C. During that year the *Iowa Placement Examinations of English-Training* were given to 109 freshmen in the Spring of 1935.

Compositions were collected for analysis. The checking of the compositions was done by the writer with very careful reference to the *Criterion*.[2] The compositions used had already been checked by the regular teachers of English. More errors were located by means of the checking done with the *Criterion* than had been found by means of the checking done without it. The study of status resolved itself into a study of the results of the testing survey, and of the analysis of compositions. A suggested program grew out of the interpretation of these results.

The data from the 1931–1932 testing program are summarized in Table II, the basic or key-table in this study.

This table reports the findings on status of proficiency in written English revealed by the test scores made on the *O'Rourke Grammar Survey Tests, Forms A and B*,[3] taken by the four groups from the three colleges; on the *Iowa Placement Examinations of English-Training, Forms X and Y*, taken by the General Group and by the Remedial Group at College C; and on the *Cook Spelling Test, Forms A and B*,[4] taken by the four groups

[1] Walter W. Cook, "The Measurement of General Spelling Ability Involving Controlled Comparisons Between Techniques." Unpublished Doctor's thesis. Iowa City, Iowa: State University of Iowa, 1932.

[2] Harry A. Greene, *A Criterion for the Course of Study in the Mechanics of Written Composition*. University of Iowa Studies, New Series, Number 4. Iowa City, Iowa: State University of Iowa, 1933.

[3] L. J. O'Rourke, *O'Rourke Survey Tests of English Usage, Forms A and B*. Washington, D. C.: Educational and Personnel Publishing Company, 1930.

[4] W. W. Cook, *op. cit.*

TABLE II

STATUS OF THE FRESHMEN TESTED AT THE THREE COLLEGES, 1931–1932*

| 1 Test | 2 Part of Test | 3 College | 4 Group | 5 Number of Cases | 6 Average Fall | 6 Average Spring | 7 Difference or Gain | 7 Sigma of the Difference | 8 S. D. Fall | 8 S. D. Spring | 9 Correlation Fall vs. Spring | 10 Sigma of the Average Fall | 10 Sigma of the Average Spring | 11 Proportional Gain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **O'Rourke Grammar Survey Tests** | Total | A | Freshmen | 54 | 58.04 | 59.48 | 1.44 | 1.01 | 10.140 | 10.550 | 0.748 0.06 | 1.38 | 1.44 | 0.034 |
| | Test 3 | | | | 11.37 | 11.22 | —0.15 | 0.47 | 3.068 | 3.515 | 0.461 0.11 | 0.42 | 0.48 | Loss |
| | Test 2 | | | | 21.35 | 22.57 | 1.22 | 0.60 | 4.926 | 4.561 | 0.565 0.09 | 0.67 | 0.62 | 0.069 |
| | Test 1 | | | | 25.11 | 25.72 | 0.61 | 0.53 | 4.263 | 4.449 | 0.603 0.09 | 0.58 | 0.61 | 0.040 |
| | Total | B | Freshmen | 62 | 57.68 | 67.81 | 10.13 | 0.91 | 12.992 | 12.974 | 0.847 0.04 | 1.65 | 1.65 | 0.239 |
| | Test 3 | | | | 10.50 | 12.66 | 2.16 | 0.50 | 3.063 | 4.036 | 0.404 0.11 | 0.39 | 0.51 | 0.227 |
| | Test 2 | | | | 20.60 | 26.08 | 5.48 | 0.53 | 6.176 | 6.254 | 0.777 0.05 | 0.78 | 0.79 | 0.282 |
| | Test 1 | | | | 26.42 | 28.95 | 2.53 | 0.51 | 5.804 | 4.930 | 0.736 0.06 | 0.74 | 0.63 | 0.186 |
| | Total | C | Freshmen (General Group) | 109 | 54.21 | 62.79 | 8.58 | 0.77 | 11.247 | 10.612 | 0.736 0.04 | 1.08 | 1.02 | 0.187 |
| | Test 3 | | | | 10.99 | 12.80 | 1.81 | 0.39 | 3.760 | 3.835 | 0.432 0.08 | 0.36 | 0.37 | 0.200 |
| | Test 2 | | | | 18.69 | 22.61 | 3.92 | 0.40 | 5.150 | 4.768 | 0.689 0.05 | 0.49 | 0.46 | 0.183 |
| | Test 1 | | | | 24.62 | 27.14 | 2.52 | 0.39 | 5.102 | 4.051 | 0.639 0.06 | 0.49 | 0.39 | 0.163 |
| | Total | C | Freshmen (Remedial Group) | 16 | 34.38 | 56.06 | 21.68 | 2.29 | 8.395 | 10.292 | 0.602 0.16 | 2.10 | 2.57 | 0.330 |
| | Test 3 | | | | 6.56 | 10.94 | 4.38 | 0.63 | 2.397 | 3.230 | 0.626 0.15 | 0.599 | 0.81 | 0.325 |
| | Test 2 | | | | 11.88 | 20.81 | 8.93 | 1.12 | 4.484 | 4.640 | 0.522 0.18 | 1.12 | 1.16 | 0.319 |
| | Test 1 | | | | 16.06 | 24.30 | 8.24 | 1.03 | 3.172 | 3.836 | 0.317 0.22 | 0.79 | 0.96 | 0.344 |
| | Total | All | Freshmen | 225 | 56.08 | 63.38 | 7.30 | 0.55 | 11.660 | 11.610 | 0.753 0.03 | 0.777 | 0.774 | 0.166 |
| | Test 3 | Three | of all but | | 10.95 | 12.38 | 1.43 | 0.37 | 3.433 | 5.879 | 0.409 0.06 | 0.228 | 0.258 | 0.158 |
| | Test 2 | Schools | the Reme- | | 19.85 | 23.56 | 3.71 | 0.29 | 5.403 | 5.527 | 0.677 0.04 | 0.360 | 0.368 | 0.184 |
| | Test 1 | | dial Group | | 25.24 | 27.22 | 1.98 | 0.27 | 5.177 | 5.011 | 0.677 0.04 | 0.345 | 0.334 | 0.079 |
| **Cook Spelling Test** | | A | Freshmen | 54 | 43.76 | 48.43 | 4.67 | 0.91 | 10.380 | 5.210 | 0.810 0.09 | 1.41 | 0.71 | 0.748 |
| | | B | | 62 | 33.71 | 47.46 | 13.75 | 0.48 | 6.840 | 4.660 | 0.860 0.06 | 0.897 | 0.61 | 0.844 |
| | | C | (G. G.) | 34 | 39.73 | 45.64 | 5.91 | 1.04 | 10.980 | 6.780 | 0.880 0.06 | 1.91 | 1.18 | 0.575 |
| | | C | (R. G.) | 15–16 | 35.37 | 44.43 | 9.06 | 1.24 | 9.046 | 5.602 | 0.887 0.05 | 2.34 | 1.45 | 0.619 |
| **Iowa Placement Examinations of English Training** | Total | C | Freshmen (General Group) | 109 | 96.59 | 119.17 | 22.58 | 1.59 | 32.461 | 33.450 | 0.802 0.03 | 3.11 | 3.20 | 0.217 |
| | Part 4 | | | | 29.31 | 32.41 | 3.10 | 0.72 | 9.370 | 8.888 | 0.657 0.05 | 0.90 | 0.85 | 0.149 |
| | Part 3 | | | | 19.24 | 25.85 | 6.61 | 1.06 | 12.415 | 11.814 | 0.618 0.06 | 1.19 | 1.13 | 0.218 |
| | Part 2 | | | | 21.12 | 33.58 | 12.46 | 1.04 | 10.444 | 10.244 | 0.482 0.07 | 1.00 | 0.98 | 0.431 |
| | Part 1 | | | | 26.75 | 27.30 | 0.55 | 0.74 | 10.651 | 13.112 | 0.814 0.03 | 1.02 | 1.26 | 0.217 |
| | Total | C | Freshmen (Remedial Group) | 16 | 37.81 | 84.50 | 46.69 | 5.39 | 6.775 | 20.524 | 0.011 0.00 | 1.69 | 5.13 | 0.288 |
| | Part 4 | | | | 19.75 | 23.13 | 3.38 | 1.00 | 6.814 | 5.290 | 0.812 0.08 | 1.70 | 1.32 | 0.111 |
| | Part 3 | | | | 1.63 | 13.13 | 11.50 | 2.79 | 4.014 | 12.226 | 0.406 0.21 | 1.00 | 3.05 | 0.237 |
| | Part 2 | | | | 3.31 | 25.86 | 22.55 | 2.26 | 4.354 | 9.096 | —0.254 0.23 | 1.09 | 2.27 | 0.482 |
| | Part 1 | | | | 13.75 | 22.38 | 8.63 | 2.60 | 6.591 | 9.571 | 0.219 0.24 | 1.65 | 2.39 | 0.238 |

* The Appendix, Part II, Sections A, B, C, D, and E, of the complete study, on file in the library of the State University of Iowa, Iowa City, Iowa, contains the frequency tables and other supporting tabulations of data from which this table is derived.

at the three colleges. One form of each test was given in the Fall, and another, in the Spring. The coefficients of correlation given in column 9 of the table indicate that very little gain had been made from the Fall test to the Spring test by the general groups. On the other hand, the low correlations for the Remedial Group indicate quite pronounced gains for that group from Fall to Spring tests. A glance at this column will show that the reliability was high in total score averages for all but the Remedial Group on the *Iowa Placement Examinations*. The small number of cases and the unusually large amount of gain may account, in part, for the low correlation for the total score averages for this group. The negative correlation on Part 2 of the same test for this group has the effect of lowering the correlation on the total.

The calculated sigma of the correlation and the sigma of the average, in almost every case, indicate that the data are reliable enough to be used as bases for drawing implications and conclusions. In the case of the *Iowa Placement Examinations* taken by the Remedial Group, the small number of students (sixteen) makes conclusions on the results for the group as a whole less reliable than those for the larger groups. However, the group will be studied both for status as a whole, and by cases, in order to provide a check on the study of status for the larger groups. Chapter V of the thesis gives a detailed report on the status of the Remedial Group.

The status for each group on all the tests at the beginning of the year, expressed in terms of the arithmetical means, is given in the sixth column of Table II, first part.

In the case of the *O'Rourke Tests,*[5] the status revealed by the average scores made on *Form A* of the test, taken in the Fall, was below the college norm for each group. The status of the Remedial Group was far below the norm.

On the *Iowa Placement Examinations,* the General Group at College C was slightly above the norm at the beginning of the year, while the Remedial Group was far below.

For the *Cook Spelling Test,*[6] a comparison of the means in spelling shows that the Remedial Group was not the lowest in initial status. The ratings are high enough for all groups to show that the spelling test used was not adequate for measuring the spelling ability of the students who were the best in spelling. (This is especially true of the final tests, on which 62 freshmen out of 234 made perfect scores. Many others made scores ranging from 47 to 49, when 50 was the perfect score. The actual gains, therefore, may be greater than the data indicate.)

Column 6, part 2, of the same table presents the status on the second form of the test, taken at the end of the year. On the *O'Rourke Tests,*[7] College B and the General Group at College C exceeded the norm for total score; College A and the Remedial Group did not quite attain it. The Remedial Group, however, made a far greater gain from Fall score to Spring score than did any other group and stood at an average of 56.06 points at the end of the year, while the norm is 67.17. The average score for the 225 freshmen of the general groups at the end of the year was 63.38 as compared with the average score of 56.06 for the sixteen members of the Remedial Group. The statistical significance of these gains is made clear in Table II.

By the end of the year, the General Group at College C had gained 22.58 points from average to average, and was well above the norm. Seventy-two cases out of this group of freshmen were still below the norm in total score on this test. The Remedial Group had gained 46.69 points, but was still 20.48 points below the norm on general average. One out of the sixteen members of the Remedial Group had exceeded the norm.

At College A, the gains on total score from Fall to Spring were very small, only 1.44 points. On Parts 1 and 2 of the test, the gains were small and on Part 3, there was a loss. The proportional amount of gain was almost negligible.

The average scores at the end of the year on the *Cook Spelling Test*[8] were so high that there seemed to be additional proof for stating that this test had not been difficult enough to measure ability above that of the average freshman.

At College B, the gains from Fall testing to Spring testing were considerably greater than were those for freshmen as made at College A, both on total scores and on scores

[5] L. J. O'Rourke, *op. cit.*
[6] W. W. Cook, *op. cit.*
[7] L. J. O'Rourke, *op. cit.*
[8] W. W. Cook, *op. cit.*

for separate parts of the *O'Rourke Tests*.[9] The proportional gains were much greater.

For College C, the same kinds of gains fell between those of the other two colleges for the General Group of freshmen. For the Remedial Group the point gains and the proportional gains were greater than for any of the groups mentioned. The gains were so much greater that they could not be accounted for by mere chance or by the single fact that the Remedial Group had twenty-four more class-hours in their first quarter than did the other students tested at College C. Their gains were greater than the gains made by the total freshman group (General Groups) of 225 students from the three colleges.

The returns from the *O'Rourke Test* were not dissimilar from those from the *Iowa Placement Examinations*. The comparison of the returns follows: The *Iowa Placement Examinations* were given at the home school only, making comparisons between the General Group of College C and the Remedial Group the only ones possible. The gains for the achievement on *Iowa Placement Examinations of English-Training* were greater on total score and on the separate parts of the test for the Remedial Group than for the General Group in terms of points. In terms of proportional gains, the Remedial Group exceeded the General Group on all but Part 4 of the test. On Part 4, the General Group had the slight advantage of .038 over the Remedial Group. (The proportional gain was found by dividing the actual gain made from Fall score to Spring score by the possible gain. The possible gain is that which could have been made from the standing on the Fall test to a perfect score.)

The one exception is that of the spelling gains made by College B, which were 4.69 points higher than those for the Remedial Group. This may be accounted for, in part, by the fact that the initial average for the Remedial Group was 1.66 points higher than that for the group from College B.

On the *Cook Spelling Tests*,[10] the greatest gain was made by College B, the next greatest by College A, the third greatest by the Remedial Group, and the fourth greatest by the General Group at College C. In the last named case, a fair comparison cannot be made, however, since the fact that some papers were lost from the Fall returns made a balanced comparison of scores for the entire number of cases impossible.

On the *O'Rourke Tests*,[11] the greatest gains made by each group were those made on Part 2, the test on adjectives, adverbs, prepositions, and conjunctions. The greatest gains on the *Iowa Placement Examinations* were made in punctuation by the General Group and by the Remedial Group.

The data mentioned from this one table, Table II, together with the statistical calculations included in the table to give evidence of reliability, indicate that the Remedial Group made greater actual point gains in almost every part of the three tests reported upon in this table than did any other group.

It seems safe to conclude that the tests indicate that very significant gains have been made on all of the tests by the Remedial Group. In terms of proportional gains (column 11), those made by the Remedial Group are about twice those made by all freshmen on the *O'Rourke Tests*. On the *Iowa Placement Examinations*, the proportional gain total for the General Group is .217, while that for the Remedial Group is .288. The difference, however, between the average gains in points made by the two groups is 24.11, in favor of the Remedial Group.

The greater gains made by the Remedial Group may be accounted for largely in terms of type of training given. The average of intelligence for the Remedial Group as compared with that for the General Group at College C in 1931–1932 was much lower. In fact, the Remedial Group ranked lower than did the freshman group at College C in 1934–1935. Data in regard to intelligence were collected.

The testing program of 1934–1935 revealed a status at School C similar to that found in the same school in 1931–1932.

Error counts were made for several sections of the tests. A study of errors made in composition writing by the Remedial Group, and by thirty of the freshmen of 1934–1935, was worked out in order to find whether the errors in basic skills in written English made on compositions were the same as those made on the tests by large percentages of the students. The types of errors on both tests and compositions were similar.

[9] L. J. O'Rourke, *op. cit.*

[10] W. W. Cook, *op. cit.*

[11] L. J. O'Rourke, *op. cit.*

Any study is beset by limitations, some of which are subjective, and some, objective. The chief limitations in the present study are those usually conceded.

The summary and the conclusions in regard to status follow. This study has tended to verify empirical knowledge in regard to the status of proficiency in the basic skills in written composition attained by college freshmen in three teachers colleges of the Middle West.

The findings in regard to status are significant, when it is remembered that the students who go out to teach with low rating on these tests will not be able to give adequate training to those whom they are to teach at the elementary school level. As a result, many more students will come to college unprepared in basic skills in written composition.

It was found, then, from the studies of 1931–1932 and 1934–1935 that:

1. The status of proficiency for the freshmen of 1931–1932, as measured by average scores on the tests used, was low at the beginning of the year for all freshmen. It was very low for those in the Remedial Group.

2. Every group of freshmen made some gain by the end of the year on all of the tests, save the freshmen from the first college. On Test 3 of the O'Rourke battery they registered a small loss.

3. The status of proficiency at the end of the year (1931–1932) was above the norm in too few of the possible cases on all the tests given. In general, it would have to be admitted that the status was somewhat low on all the tests, especially when it is remembered that many of the students tested were going out to teach the subjects in which they had made low scores.

4. The status for freshmen, sophomores, juniors, and seniors of 1931–1932 was not dissimilar from that of the freshmen, if that conclusion may be inferred from the results for the small number of these classes tested.

5. The greatest gains on the tests were made by the Remedial Group. The data supporting this statement and the details of the program worked out would seem to point to the fact that the program of training given the Remedial Group was somewhat effective. From this program certain procedures have been selected as worthy of further use in an experimental way. These procedures appear in Chapter VII of the writer's thesis.

6. The points in English which large percentages of these students missed on the tests and in the compositions are those which should have been taught in the elementary school, according to Cesander's[12] study of grade placement.

7. The types of errors appearing most frequently on the tests are those which appeared most frequently in the compositions analyzed. This fact indicates that the use of these tests for purposes of diagnosing the needs of students in the fields tested is a valid procedure.

8. The status of proficiency on the part of freshmen at College C in the Spring of 1935 was about the same as that for the freshmen of the same school in the Spring of 1932, judging from results on comparable forms of the *Iowa Placement Examinations of English-Training*. It seems somewhat evident that the program of instruction in English composition at that school has been static for the period of years from 1931 to 1934. Intelligence test scores, for both 1931 and 1934, bear out the fact that native ability was about on a par for the groups of freshmen tested during those two years. Therefore, the condition cannot be attributed to variations in native ability.

In general, it may be concluded from the study of status made at the given institutions through the media of testing and composition analysis that only a small number of students are adequately prepared for the high art of teaching the basic skills in written English. It is a problem which demands serious attention and action.

[12] Paul K. Cesander, "A Study of Pupil Usage as a Factor in the Grade Placement of Certain Items of Punctuation." Vol. I. Unpublished Doctor's thesis. Iowa City, Iowa: State University of Iowa, 1931.

# THE EVALUATION OF A TECHNIQUE FOR MEASURING IMPROVEMENT IN ORAL COMPOSITION*

R. F. NETZER
*Arizona State Teachers College, Tempe*

During the past decade workers in the elementary school field have pointed out that emphasis on the development of the skills in written and oral expression must coincide more closely with the life demands on language. They have shown that most of the out-of-school demands on language ability are of the oral type and that teachers spend too much time on the mechanical and technical aspects of language. Perhaps the relative ease of judging the mechanical and technical aspects and the lack of an objective means of judging oral productions have led to this overemphasis on written language. Also, textbooks and courses of study have been much less provident of materials for oral language.

Modern curricular emphasis is being adjusted to provide more adequate oral language instruction and with this expansion have come certain problems in materials and methods. One of these problems is the discovery of a technique for measuring improvement in oral composition. A common means of measuring improvement in written composition has been a scale. Whether this technique is applicable to the measurement of improvement in oral composition has not been determined, because until very recently no method of making exact reproductions of oral compositions was available. Now that exact reproductions can be made by means of recording equipment it is possible to evaluate the scale as a method of measuring improvement in oral composition. The purpose of this study was to evaluate experimentally an oral language scale as a means of revealing group and individual growth in ability to use oral language skills. Because no adequate oral language scales were available, the construction of such instruments became a necessary preliminary to the experimental phases of the study.

In this discussion the term "scale" is considered to mean a group of children's productions ranging from very poor to very good, arranged in ascending order and of such quality that the second is approximately as much better than the first as the third is better than the second and as the fourth is better than the third.

Research has shown that rating by a general merit scale is much more reliable than rating by an analytical type of scale, that themes on the same topic and themes with a wide range of quality are easier to rate, and that there is a definite relationship between general quality rating and rating for specific factors. The inadequacy of written scales for judging oral compositions has also been demonstrated, since they do not show the influence of difference in vocabulary and expression found in oral composition. Children, like adults, use words and expressions in speaking which they do not employ in writing. In a previous investigation the writer found it impossible to use a written composition scale to rate oral compositions. Too small a difference in quality between the poorest and best production was revealed through the use of a written scale. The sentence structure and usage in oral composition are inferior to those of written composition. Some of the constructions which are considered unpolished when they appear in written form, apparently add to the effectiveness of oral production—sentence fragments, for instance. It is also valid to say that the vocabulary of oral composition is not so limited as the vocabulary of written composition, since children and adults often use words in speaking which they do not employ in writing because of inability to spell such words. Thus the standards for evaluating oral and written composition must differ.

The first problem was to find satisfactory means of stimulating oral composition. After trying out a large number of procedures which had been suggested by teachers and supervisors, three, which appeared satisfactory, were selected for use. They were a picture, an incomplete story, and an object.

* Summary of a doctoral dissertation, University of Iowa, 1937.

The next problem was to select suitable materials of each of these three types. The results of studies pertaining to the preference of children for certain types of pictures do not agree and there have been no studies concerning the relative ability of different types of pictures to stimulate language expression. Likewise there is no reason to believe that the picture which is most liked will best stimulate expression. Beautiful pictures and scenes may be much liked but be beyond the power of anyone but an artist to express in words. It is reasonable to assume that the picture selected must contain a few significant details which are prominent and in the foreground.

The writer has been unable to discover any studies concerning the type of story which best stimulates language expression. It seems reasonable that the incomplete story selected must tell enough to give the main idea but still leave something to be told, must end at an interesting point so that the child will be stimulated to complete it, and must not exactly determine the completion but leave more than one solution possible.

Studies of the relative ability of objects to stimulate expression are lacking, but it is reasonable to require that the object selected be something that a child has seen or read about, realistic in shape and color, and perhaps something a child can handle. Each of the things selected must be within the range of experiences and understanding of children and must be something which allows for extensive expression by those capable of expressing themselves in greater length and detail.

The three stimuli selected for each grade on the basis of the above criteria were presented to groups of pupils in the fourth, fifth, and sixth grades of the University of Iowa Elementary School and the pupil's responses were recorded by the Iowa Oral Language Recording Equipment. The children were told that the investigator was making a collection of children's stories and that their responses were to be recorded. The children were not "microphone conscious" due to the fact that they had seen the microphones before and had had their responses recorded for other studies in previous years. When the hour was over many of them wanted to remain longer and, when told that this was impossible, asked if they might return at another time.

The records of the 135 responses were transcribed and proofread. Fifty-one were from the fourth grade, forty-two from the fifth grade, and forty-two from the sixth grade. No effort was made to limit the number of compositions. All that were given were recorded and transcribed. The compositions varied in length from thirty-eight to 1014 words.

An analysis of these compositions showed that in general the object used was the best of the three methods of stimulating oral composition, the story second best, and the picture the poorest method, although the differences were not statistically significant. On the basis of these results it seemed advisable to use the same three types of stimuli and to select one of each type for each of grades two and three. Those selected were considered suitable for the purpose by several teachers of grades two and three.

Permission was obtained to record the responses of children in four school systems in Iowa. These systems ranged in size from a small town to one of the larger cities of the state. Because the recording equipment required 110 A. C. 60 cycle current it was impossible to use all the school systems which were willing to cooperate. The teachers in these systems were requested to select the five poorest and five best pupils in oral language in their rooms. The writer was primarily interested in obtaining samples which ranged from extremely good to extremely poor rather than in the grade level from which they were obtained. This range was necessary in order to obtain adequate spread for the scale. Furthermore, samples with a wide range of quality were found to be more easily and reliably rated.

The recording equipment was taken to each school, the stimuli presented, and the pupil's responses recorded in the same way as at the University Elementary School. The responses were transcribed and proofread. In this way 249 responses were obtained which ranged from those of poor second graders to those of good sixth graders. These responses were combined with those obtained at the University Elementary School. This group of 384 responses was used as the source of material for developing the scales.

In deriving the quality scales an exact copy of each of the 384 responses was used. Although the compositions contained many incomplete sentences, sentence fragments, and repetitions of words and phrases, they were not smoothed because the exact response was desired for analysis.

Because it would be nearly impossible and extremely unreliable to have the judges arrange the total 384 responses in order of quality from poorest to best, it was decided that some preliminary elimination was necessary. This was accomplished by having each of five qualified judges independently arrange this total group of responses in ten piles, each pile representing a different level of quality from very poor to very good. Thus, those that each judge considered the best he placed in pile one, those considered poorest in pile ten, and those judged to be of a value between these two extremes he placed in the appropriate one of the eight intervening piles.

Those most often falling in each of the ten groups were selected for use in determining the final scales. These were consciously selected so that they would represent the total range of quality from very inferior to very superior. For example, samples 31 and 47 were selected as most representative of the third group because three out of five judges placed sample 31 in group three and two placed it in group two, and thus it probably represents a value slightly less than three; sample 47 was placed in group three by four judges and in group four by one judge, and thus it represents a value slightly more than three. In this manner twenty-five samples were selected so that they represented as nearly as possible a continuous range from good to poor.

The original plan was to develop a general scale from the responses to the three stimuli. This was abandoned because the different stimuli influenced to some extent the responses, i.e., the response to an incomplete story differed greatly from the response to a picture because of the inherent difference in the stimulus itself. In this way, also, a frequent criticism of written scales (that the responses are of a single general type and difficult to make comparisons with unless the same stimulus is used) was overcome. Therefore twenty-five samples resulting from the picture as a stimulus, twenty-five samples resulting from the object as a stimulus, and twenty-five samples resulting from the story as a stimulus were selected as the basis of the three types of scales.

The samples for each stimulus were scaled by means of the method developed by Thurstone.[1] The basis of this technique is the composite opinion of qualified judges who rated each of the samples in relation to every other sample. Each of the samples was rated by twenty-five judges. A chance-half correlation was calculated for these judgments to obtain some idea of the consistency of judgments. The correlations which resulted were .98, .98, and .97 for the judgments on the story, picture, and object samples respectively. This indicated that there was close agreement on the part of the judges as to the relative position given each sample.

The evaluation of the technique was made by using the scales to assign values to the oral responses of each pupil. Almost exactly a year after the first oral language records were obtained the same schools were revisited and additional recordings of the oral productions of the same pupils were made. Each pupil was asked to respond to the stimuli to which he had responded the year before. These responses were recorded in the same way and under the same conditions as those of the previous year. Two sets of records, a year apart, were then available for analysis.

Three judges were trained to use these scales. When a judge was able to use the scales with facility and accuracy he was given copies of the 1936 samples, arranged according to stimulus, and a copy of each of the three scales. He worked independently, assigned a value to a sample, and recorded these ratings on a sheet prepared for the purpose. After a group was completed the judge was allowed to go back and adjust the values assigned if he wished. Several weeks later each judge again practiced using the scales and was then given the 1937 samples to be rated in the same way.

These ratings were used to evaluate the scales. There were 249 samples recorded in 1936 but in 1937 it was possible to locate only 199 of the same children. The others had transferred from the school systems in which they had formerly been registered. The evaluation of the scales is therefore based on the 199 pairs of responses. The results of

[1] L. L. Thurstone, "An Experimental Study of Nationality Preferences," *Journal of General Psychology*, I (July, 1928), 405–423.

these ratings are summarized in Table I which shows that with one exception the differences are statistically significant.

For the purpose of analyzing individual gains and losses in oral composition ability, as measured by these scales, distributions were made of the number and percentage of pupils who made a gain or a loss for each stimulus from 1936 to 1937. Table II gives a summary of the results. This table shows that the gain in oral language ability, as measured by the scales, was always greater than the loss after the lapse of a year. When the story was used as a stimulus and the resulting responses were rated by means of the story scale, 22 percent more pupils made a gain than made a loss. Similar analysis shows that when the picture was used the excess of pupils making gains rather than losses was 40 percent and when the object was used this excess gain was 54 percent. Table II also shows that with one exception the percentage of pupils making a gain exceeded the percentage of pupils making a loss for each interval when the gains and losses were tabulated by intervals of one. That is, more pupils made a gain of from zero to one unit than made a loss of from zero to one unit; more made a gain of from one to two units than made a loss of from one to two units, and similarly for from two to three units and above three units. The only exception that was found was when the story was used as a stimulus. Two percent more pupils made a loss of over three units than made a gain of over three units for this stimulus. It was pointed out before that when this scale and stimulus were used the amount of improvement found was not statistically significant. However, this stimulus did show a greater percentage of pupils gaining than losing for the total group and in the intervals from one to three.

Table II also shows that, with the exception of the responses to the story stimulus, the average gain of those who gained was always greater than the average loss of those who lost. The excess of average gains over average losses was —.1, .6, .6, and .4 for the story, picture, object, and the combination of the three respectively. Previous to this time it has been taken for granted that most children would make a gain in oral language ability during a year's instruction in this subject. However, adequate proof of this fact and a definite knowledge of the extent of this gain were not available. This experiment furnishes proof that this assumption of a significant growth in oral language ability is correct. It shows that as a result of a year's school experience the majority of the pupils made significant gains in ability to express themselves orally.

As a result of this study it seems that the following conclusions are valid:

1. The experience in this study indicates that the Thurstone method of scale construction is a reasonably effective means of scaling children's oral productions.

TABLE I

SUMMARY OF THE RESULTS OF THE RATINGS ON THE 1936 AND 1937 SAMPLES

| | A. M. | S. D. | S. E. AM | r | Diff. | S. E. diff | C. R. |
|---|---|---|---|---|---|---|---|
| Combined | | | | | | | |
| 1936 | 3.35 | 2.13 | .15 | .62 ± .03 | .76 | .13 | 5.85 |
| 1937 | 4.11 | 1.96 | .14 | | | | |
| Story | | | | | | | |
| 1936 | 3.86 | 2.32 | .33 | .56 ± .07 | .30 | .28 | 1.07 |
| 1937 | 4.16 | 1.62 | .23 | | | | |
| Picture | | | | | | | |
| 1936 | 2.58 | 1.84 | .20 | .58 ± .05 | .77 | .17 | 4.53 |
| 1937 | 3.35 | 1.68 | .18 | | | | |
| Object | | | | | | | |
| 1936 | 3.97 | 2.01 | .25 | .62 ± .05 | 1.08 | .22 | 4.91 |
| 1937 | 5.05 | 2.12 | .26 | | | | |

TABLE II

PERCENTAGE OF PUPILS MAKING A GAIN OR LOSS FROM 1936 TO 1937

| | Gain per Unit | | Loss per Unit | |
|---|---|---|---|---|
| Story | | | | |
| Gain 60% | 0 to 1 | 32% | 0 to —1 | 20% |
| Loss 38% | 1.2 to 2 | 12% | —1.2 to —2 | 6% |
| Average gain 1.4 | 2.2 to 3 | 12% | —2.2 to —3 | 6% |
| Average loss 1.5 | 3.2 up | 4% | —3.2 down | 6% |
| Picture | | | | |
| Gain 69% | 0 to 1 | 25% | 0 to —1 | 18% |
| Loss 29% | 1.2 to 2 | 23% | —1.2 to —2 | 7% |
| Neither 1% | 2.2 to 3 | 15% | —2.2 to —3 | 3% |
| Average gain 1.7 | 3.2 up | 6% | —3.2 down | 1% |
| Average loss 1.1 | | | | |
| Object | | | | |
| Gain 79% | 0 to 1 | 32% | 0 to —1 | 13% |
| Loss 25% | 1.2 to 2 | 19% | —1.2 to —2 | 8% |
| Neither 2% | 2.2 to 3 | 14% | —2.2 to —3 | 2% |
| Average gain 1.7 | 3.2 up | 14% | —3.2 down | 2% |
| Average loss 1.1 | | | | |
| Combination of Story, Picture and Object | | | | |
| Gain 68.5% | 0 to 1 | 28.5% | 0 to —1 | 17% |
| Loss 30% | 1.2 to 2 | 18.5% | —1.2 to —2 | 7% |
| Neither 1% | 2.2 to 3 | 13.5% | —2.2 to —3 | 3.5% |
| Average gain 1.6 | 3.2 up | 8 % | —3.2 down | 2.5% |
| Average loss 1.2 | | | | |

2. From the point of view of English, improvement in oral composition can be measured effectively by means of an oral composition scale.
3. Two of the three scales developed in this study prove to be useful for measuring improvement in children's oral productions. The critical ratios for the picture and object scales are 4.53 and 4.91 respectively.
4. Even though the critical ratio for the story scale is only 1.07, the story scale would possibly be useful for measuring improvement, if the inherent limitations of the responses due to the type of stimulus used are taken into account during the rating process.
5. Observation of the work of the judges during the training period indicates that teachers can be trained to use an oral composition scale to judge the oral productions of their pupils with a resultant increase in the reliability of their recorded judgments.
6. As would be expected, more pupils increase in ability to express themselves orally as a result of a year's school experience than decrease in this ability, and the average increase of children in ability to express themselves orally is greater than the average decrease in ability to do so.

# THE RELATIONSHIP BETWEEN SCORES ON THE SCHOLASTIC APTITUDE TEST AND COLLEGE MARKS IN CHEMISTRY

M. Richard Dickter
*Philadelphia, Pennsylvania*

## The Scholastic Aptitude Test

In 1926 the College Entrance Examination Board devised an aptitude test, known as the Scholastic Aptitude Test. It is a test offered to colleges and universities for the purpose of testing all candidates for admission to these institutions. No institution, however, is obligated to give the test; acceptance is entirely on a voluntary basis.

The test, from 1930 until 1936, the period covered by this investigation, consisted of two sections: one of which is material essentially linguistic or verbal in nature and another of which is material essentially mathematical in nature. The test aims to measure potential ability to do work on the college level along specific lines, rather than to measure ability in general. A primary purpose of the Scholastic Aptitude Test is its use as an aid in guidance and teaching in college; for example, selecting students for admission to the various fields of study, guiding them in their choice of majors, forestalling through supervision the possible failures of doubtful students, diagnosing the mediocre performance or failures of promising students, and sectioning groups for instructional purposes.[1]

### *The Verbal Section*

The Scholastic Aptitude Test was first given to applicants for college entrance by the College Entrance Examination Board in the year 1926. From 1926 to 1929 inclusive, this test consisted only of a verbal or non-mathematical section, the composition of which changed during these years. In 1926 and 1927 the verbal section consisted of nine sub-tests. In 1928 the number of sub-tests was reduced to seven, and further reduced to six in 1929. From 1930 to 1936, the period covered by this study, the verbal section consisted of three sub-tests, totalling two hundred items:

1. Antonyms
2. Double definitions
3. Paragraph reading

### *The Mathematical Section*

Not long after the inception of the Scholastic Aptitude Test, it became evident that the test was better adapted to selecting students for those courses in which verbal abilities are required, such as history, English, and Latin, than to selecting students for the mathematical and scientific courses, where mathematical abilities are essential. It was felt that the verbal section alone was not sufficient, because it measured linguistic ability only and not the ability that is thought of as the mathematical and scientific.[2]

The dissatisfaction became sufficiently acute to move the Commission on Scholastic Aptitude Tests of the College Entrance Examination Board to add a second section, in 1930, to the Scholastic Aptitude Test, after experimenting with such a section in 1928 and 1929. This second part was a mathematical section consisting of one hundred items, arranged in the order of difficulty, dealing with arithmetic, algebra, and geometry, and designed better to select students for the mathematical and scientific courses.

## The Problem

The Scholastic Aptitude Test is an instrument furnished by the College Entrance Examination Board and is used by colleges and universities as an aid in selecting students for admission to the various fields of study, and as an aid in guidance. Considerable time and effort have been, and are being, spent on it. The question arose at the University of Pennsylvania of the degree of relationship between the scores made by students on the mathematical section of the Scholastic Apti-

[1] E. G. Bill and H. Pennypacker, "Two Deans Consider the Scholastic Aptitude Tests," *Independent Education*, III (October, 1929), 17.

[2] *Annual Report of the Commission on Scholastic Aptitude Tests*, 1929, p. 38.

tude Test at the time of their entrance to the University and their latter academic success, as judged by their marks, in specific courses in first-year chemistry. This study was undertaken in order to answer this question.

At the same time, the relationship between the scores made by the same students on the verbal section and their marks in the same courses was also determined, thus making it possible, in addition, to compare the relative merits of the two sections as far as first-year chemistry is concerned. These relationships were investigated for the period 1930–36, the years during which both sections of the Scholastic Aptitude Test were given.

The situation is an individual one in each university, so that results obtained at one university are not necessarily directly comparable with similar results obtained at other universities. Yet, the findings at the University of Pennsylvania, in the sense that they indicate trends, will be of value to other interested groups outside this university.

## The Courses Studied

Two one-semester chemistry courses, generally studied in the freshman year, were included: Chemistry 1 (General Inorganic Chemistry) and Chemistry 3 (Qualitative Analysis). Chemistry 1 is a prerequisite for Chemistry 3.

## The Data

All the data for this investigation were obtained directly from the records of the individual students. Likewise, as far as the University of Pennsylvania is concerned, the data do not represent a sampling—all students who came within the scope of the investigation, that is, all students who had been admitted to the University of Pennsylvania during the years from September, 1930, to February, 1936, and who, up to June, 1936, had taken one or both of the two chemistry courses studied, were included.

In order to obtain the data, all the undergraduate schools[3] of the University were visited and the records of their former and present students examined. When the record of a student was found who had been admitted during the period between the academic years 1930–31 and 1935–36 and who had taken one or both of the two chemistry courses, the name of the student, his school, whether he had studied Chemistry 1 or both Chemistry 1 and Chemistry 3, and his final mark in each of these courses were noted. Then the admission slips of these students were located and their scores on the verbal and mathematical sections of the Scholastic Aptitude Test recorded.

In this way, a total of 2603 cases, or marks, were obtained for investigation, of which 1742 were in Chemistry 1 and 861 in Chemistry 3. About fifty per cent of the cases in Chemistry 1 came from The College, about thirty-five per cent from The Towne Scientific School, with the remainder distributed among the other schools; whereas about sixty-five per cent of the cases in Chemistry 3 came from The College, about twenty-five per cent from The Towne Scientific School, with the remainder distributed among the other schools.

## Statistical Treatment

The relationship between test scores and course marks was determined by calculating the coefficients of correlation between the variables involved by means of the Pearson product-moment method. Coefficients of correlation were computed between the scores on the verbal and mathematical sections of students admitted to the University of Pennsylvania in each of the six academic years covered in this investigation, 1930–31 to 1935–36, inclusive, and the marks they later made in the courses under consideration. Four such correlations were made for each year—two between verbal test scores and marks in each of the courses and two between the mathematical test scores and marks in each of the same courses (verbal test scores and mathematical test scores refer, respectively, to the scores on the verbal and mathematical sections of the Scholastic Aptitude Test).

A similar set of four correlations was calculated for the entire period 1930–36, regardless of the year of admission. Separate correlations by years were made in addition to those for the entire period because it was felt that the latter, while valuable as a summary of what had happened, would not disclose any trends or variations that might have occurred during the years.

[3] The College (for men), The College of Liberal Arts for Women, The School of Fine Arts, The Moore School of Electrical Engineering, The School of Education, The Towne Scientific School (Chemistry, Chemical, Civil, and Mechanical Engineering), and The Wharton School of Finance and Commerce.

It will be noted, likewise, that separate correlations were calculated between test scores and marks in each of the two chemistry courses, rather than between test scores and the average marks of the two courses. The latter procedure would have yielded a confused picture of the relationships sought. A relationship determined for the average of two courses could not very well be applied to either of the two courses included in the average.

The number of cases that were available for each of the correlations by years and for the entire period may all be determined from Table I. The number of cases for a course in a particular year may be obtained from the column for that year, and for the entire period 1930–36 from the last column on the right.

## Results and Conclusions

The results obtained in this investigation are presented in Tables II–VII, inclusive. Tables II, III, and IV contain the detailed results by years and Tables V, VI, and VII the corresponding summary results for the entire period 1930–36. Tables II and V show the coefficients of correlation between marks in each of the courses and scores on the verbal and mathematical sections, Tables III and VI the means of the course marks and the means of the scores on the verbal and mathematical sections, and Tables IV and VII the standard deviations of the course marks and the standard deviations of the scores on the verbal and mathematical sections.

Tables II–VII, inclusive, contain certain abbreviations the meanings of which are necessary in order to understand the tables.

### Table I

Distribution of Cases by Courses in Each Academic Year for the Entire Period 1930–36

| Course | 1930–31 | 1931–32 | 1932–33 | 1933–34 | 1934–35 | 1935–36 | Total |
|---|---|---|---|---|---|---|---|
| C. 1 | 261 | 299 | 302 | 324 | 293 | 263 | 1742 |
| C. 3 | 127 | 152 | 171 | 183 | 168 | 60 | 861 |
| Total | 388 | 451 | 473 | 507 | 461 | 323 | 2603 |

### Table II

Correlations Between Marks in the Chemistry Courses and Scores on the Verbal and Mathematical Sections of the Scholastic Aptitude Test for Each Academic Year in the Period 1930–36

| Course | 1930–31 | | | | 1931–32 | | | | 1932–33 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | V. S. | | M. S. | | V. S. | | M. S. | | V. S. | | M. S. | |
| | r | P. E. | r | P. E. | r | P. E. | r | P. E. | r | P. E. | r | P. E. |
| C. 1 | .302 | .038 | .526 | .030 | .236 | .037 | .422 | .032 | .221 | .037 | .381 | .033 |
| C. 3 | .224 | .056 | .415 | .049 | .193 | .053 | .309 | .050 | .181 | .050 | .283 | .047 |

| Course | 1933–34 | | | | 1934–35 | | | | 1935–36 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | V. S. | | M. S. | | V. S. | | M. S. | | V. S. | | M. S. | |
| | r | P .E. | r | P. E. | r | P. E. | r | P. E. | r | P. E. | r | P. E. |
| C. 1 | .362 | .032 | .458 | .030 | .314 | .036 | .443 | .031 | .107 | .041 | .366 | .036 |
| C. 3 | .256 | .047 | .329 | .044 | .158 | .051 | .267 | .048 | —.042 | .087 | .389 | .074 |

The meanings of these abbreviations are given here:

C = chemistry
V.S. = verbal section
M.S. = mathematical section
r = coefficient of correlation
P.E. = probable error of the coefficient of correlation
$X'$ = mean of the course marks
$Y'$ = mean of the scores on the verbal section or the mathematical section of the Scholastic Aptitude Test
$\sigma_x$ = standard deviation of the course marks
$\sigma_y$ = standard deviation of the scores on the verbal section or the mathematical section of the Scholastic Aptitude Test

TABLE III

MEANS OF THE COURSE MARKS AND THE MEANS OF THE SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR EACH ACADEMIC YEAR IN THE PERIOD 1930–36

| Course | 1930–31 | | | 1931–32 | | | 1932–33 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $X'$ | V. S. $Y'$ | M. S. $Y'$ | $X'$ | V. S. $Y'$ | M. S. $Y'$ | $X'$ | V. S. $Y'$ | M. S. $Y'$ |
| C. 1 | 4.2 | 45.0 | 43.5 | 4.3 | 46.5 | 45.0 | 5.0 | 43.0 | 39.5 |
| C. 3 | 4.8 | 49.0 | 46.5 | 4.7 | 50.5 | 45.5 | 4.9 | 45.5 | 41.0 |

| Course | 1933–34 | | | 1934–35 | | | 1935–36 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $X'$ | V. S. $Y'$ | M. S. $Y'$ | $X'$ | V. S. $Y'$ | M. S. $Y'$ | $X'$ | V. S. $Y'$ | M. S. $Y'$ |
| C. 1 | 4.8 | 41.5 | 40.0 | 4.6 | 42.5 | 41.5 | 4.6 | 42.5 | 43.0 |
| C. 3 | 4.9 | 41.5 | 41.0 | 4.9 | 46.5 | 44.5 | 4.5 | 45.0 | 51.5 |

TABLE IV

STANDARD DEVIATIONS OF THE COURSE MARKS AND THE STANDARD DEVIATIONS OF THE SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR EACH ACADEMIC YEAR IN THE PERIOD 1930–36

| Course | 1930–31 | | | 1931–32 | | | 1932–33 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_x$ | V. S. $\sigma_y$ | M. S. $\sigma_y$ | $\sigma_x$ | V. S. $\sigma_y$ | M. S. $\sigma_y$ | $\sigma_x$ | V. S. $\sigma_y$ | M. S. $\sigma_y$ |
| C. 1 | 2.74 | 29.30 | 30.00 | 2.67 | 27.00 | 28.55 | 2.44 | 27.80 | 29.85 |
| C. 3 | 2.82 | 29.20 | 30.70 | 2.72 | 26.05 | 28.20 | 2.76 | 27.80 | 29.55 |

| Course | 1933–34 | | | 1934–35 | | | 1935–36 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_x$ | V. S. $\sigma_y$ | M. S. $\sigma_y$ | $\sigma_x$ | V. S. $\sigma_y$ | M. S. $\sigma_y$ | $\sigma_x$ | V. S. $\sigma_y$ | M. S. $\sigma_y$ |
| C. 1 | 2.64 | 28.40 | 29.20 | 2.67 | 27.45 | 27.75 | 2.61 | 28.05 | 27.30 |
| C. 3 | 2.68 | 28.75 | 28.90 | 2.54 | 27.30 | 28.90 | 2.39 | 27.55 | 29.15 |

TABLE V

SUMMARY OF CORRELATIONS BETWEEN MARKS IN THE CHEMISTRY COURSES AND SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR THE ENTIRE PERIOD 1930–36

| Course | Verbal section | | Mathematical section | |
|---|---|---|---|---|
| | r | P. E. | r | P. E. |
| C. 1 | .253 | .015 | .423 | .013 |
| C. 3 | .189 | .022 | .315 | .021 |

TABLE VI

SUMMARY OF THE MEANS OF THE COURSE MARKS AND THE MEANS OF THE SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR THE ENTIRE PERIOD 1930–36

| Course | X′ | V. S. Y′ | M. S. Y′ |
|---|---|---|---|
| C. 1 | 4.6 | 43.5 | 42.0 |
| C. 3 | 4.8 | 46.0 | 44.0 |

TABLE VII

SUMMARY OF THE STANDARD DEVIATIONS OF THE COURSE MARKS AND THE STANDARD DEVIATIONS OF THE SCORES ON THE VERBAL AND MATHEMATICAL SECTIONS OF THE SCHOLASTIC APTITUDE TEST FOR THE ENTIRE PERIOD 1930–36

| Course | $\sigma_x$ | V. S. $\sigma_y$ | M. S. $\sigma_y$ |
|---|---|---|---|
| C. 1 | 2.65 | 28.05 | 28.85 |
| C. 3 | 2.70 | 28.00 | 29.35 |

*Analysis by Years*—See Table II

1. The correlations on the mathematical section are higher than those on the verbal section in both Chemistry 1 and Chemistry 3 each year.

2. The correlations are lower for Chemistry 3 than for Chemistry 1 in every case on both the verbal and the mathematical sections, with the exception of the mathematical section for the 1935–36 group.

*Analysis over the Entire Period 1930–36*—See Table V

1. The correlations on the mathematical section are higher than those on the verbal section for both courses, when considering results over the entire six-year period.

To determine the reliability of the difference between the obtained coefficients of correlation for the verbal and mathematical sections, the ratio of the difference of the two coefficients to the probable error of their difference[4] was calculated for each course. In both Chemistry 1 and Chemistry 3, the ratios so found were greater than four, indicating that there is complete certainty (100 chances in 100) that the obtained difference represents a true difference greater than zero.[5]

2. The correlations on both the mathematical and verbal sections are lower for Chemistry 3 than for Chemistry 1 in every year (with the exception of the mathematical section for the 1935–36 group) and for the entire six-year period. The lower correlation in the second semester of the course is probably due to the greater homogeneity of the students in the second half. If the means of the scores of the students on the mathematical section in Table VI are compared for Chemistry 1 and Chemistry 3, and similarly for the verbal section, it is seen that these means are higher for Chemistry 3 than for Chemistry 1, thus showing that in the second half of the course the students represent a group higher in those abilities tested by the Scholastic Aptitude Test, due to the elimination of the students who failed at the end of the first half.

3. The correlations on the verbal section are low for both Chemistry 1 and Chemistry 3, but those on the mathematical section are much better. It must be remembered that we are dealing with a single instrument and that we are on the college level, where the student personnel is more homogeneous than on the secondary level. The obtained coefficients of correlation are also lower than they otherwise would be, because not all the candidates who apply for admission and take the Scholastic Aptitude Test are admitted, the poorer ones being rejected.[6] The more

[4] $\frac{D}{PE\ (\text{diff. } r_1 - r_2)}$

[5] Henry E. Garrett, *Statistics in Psychology and Education*, p. 135, Table XV. New York: Longmans, Green and Co., 1926.

[6] *Twenty-Sixth Annual Report of the College Entrance Examination Board*, 1926, p. 8.

homogeneous the group, the greater is the restriction in the range of ability, and this restriction tends to lower the coefficients of correlation.[7]

4. Beginning with the June, 1936, Scholastic Aptitude Test, taken in the main by candidates for admission in September, 1936, the College Entrance Examination Board dropped the mathematical section. This move by the Board did not affect any of the students considered in this study, since the study concerns itself with students admitted between September, 1930, and February, 1936, during which time the verbal and mathematical sections did not change and were taken by all applicants.

In place of the mathematical section, the Board is offering a new series of mathematical examinations, called Alpha, Beta, and Gamma.[8] Since the Board cannot compel the acceptance of its offerings, it remains for the individual university to decide upon their acceptance or rejection.

The results obtained on the mathematical section in this investigation show sufficient promise to indicate the desirability of the continued use of such a section in the Scholastic Aptitude Test for the guidance and teaching of students in courses in chemistry. Since the original mathematical section is no longer available, and since the College Entrance Examination Board is now offering the new mathematical examinations (Alpha, Beta, and Gamma) in its place, it is suggested that these new examinations be given a trial. It would remain, however, for a future study similar to this one, by comparing the results of such a study with those found here, to determine whether the new mathematical examinations are more closely related to marks in chemistry than was the original mathematical section.

[7] Ruth K. Byrns, "Scholastic Aptitude and Freshman Achievement," *School and Society*, XXXV (May 21, 1932), 718.
A. B. Crawford, "Forecasting Freshman Achievement," *School and Society*, XXXI (January 25, 1930), 129.
Frank S. Freeman, "Predicting Academic Survival," *Journal of Educational Research*, XXIII (February, 1931), 116.
David Segel, *Prediction of Success in College*, p. 19. U. S. Office of Education Bulletin, 1934, No. 15. Washington: Office of Education.

[8] "Report of the Commission on Examinations in Mathematics," *Mathematics Teacher*, XXVIII (March, 1935), 154–66.

# A STUDY OF AURAL LEARNING WITH AND WITHOUT THE SPEAKER PRESENT[1]

J. Edwin Loder
*Principal of Havelock High School*
*Lincoln, Nebraska*

## Introduction

The quantity and quality of aural learning present problems wherever there is a radio and learning is the primary objective. Sound equipment is common in schools and yet few of the problems concerning its efficient use have been solved. As a matter of fact, with the speaker in the room no universally accepted best method of obtaining the greatest quantity and quality of aural learning has ever been established. Aural learning problems, then, have been multiplied by the introduction of loud speaker presentations.

The researches in the field represented by these problems have helped in only a small way to untangle and solve them. After studying the work which has been done with the problems that could be solved in the field, as presented by W. W. Charters in his pamphlet, *Research Problems in Radio Education*,[2] one is overwhelmed with the lack of information. More facts are needed concerning the learning of information over the loud speaker in order to establish scientific foundations for all its predictions.[3]

A great many of these facts fall directly in the field of psychology. What facts are needed, the gathering of these facts, and their verification are a challenge to those interested in psychological research. This problem was suggested to the writer by a statement made by Professor D. A. Worcester[4] in the closing of a paper read to the 1932 National Institute For Education By Radio. In connection with his report on certain investigations he made this statement, "I do not know whether or not the scores on the various materials reported in the paper would have been better had the participants been listening to a speaker in the room rather than over the radio."

[1] The complete study is on file in the office of Dr. D. A. Worcester, Teachers College, University of Nebraska.

[2] W. W. Charters, *Research Problems in Radio Education*. Information Series No. 4. New York: National Advisory Council on Radio in Education, 1932.

[3] L. Tyson and W. J. Donovan, *Retrospect and Forecast in Radio Education*. Chicago: University of Chicago Press. Pp. 254.

[4] D. A. Worcester, "What People Remember from Radio Programs," *Education on the Air*, 1932, p. 228. Columbus, Ohio: Ohio State University Press.

## The Problem

This study is concerned with the learning and retention of factual material presented over the loud speaker compared with the learning and retention of factual material presented directly by the speaker himself. Does it make a difference in the amount learned and retained whether or not the speaker is in the learner's presence?

## The Procedure

### *The Plan*

Four lessons were taught by two types of presentation and a comparison made of the learning. Four hundred and forty-nine 8th grade pupils in seven junior high schools in the Public Schools of Lincoln, Nebraska, were the subjects in this experiment. Their scores on experimental tests were studied in the light of common items of information concerning them on their school census sheets, and other information thought to be pertinent.

After a pre-test was given, four lessons were presented to the pupils, upon which they were tested immediately after each lesson and at two later dates. In presenting each lesson in each experimental set-up there were two groups of pupils. One group heard the lesson directly from the speaker with the speaker in their presence, while the other group heard the presentation simultaneously in another room over a loud speaker.

### *Tests*

All tests were of the true-false type. The first or immediate retention tests, given immediately after each lesson, were constructed first, and the questions were based directly on the factual material in each lesson. The pre-test and the second and third retention tests were the same except for the directions and

certain questions which gathered information concerning a radio in the pupil's home and concerning the reading of related information. Twenty-five questions were taken from each of the four immediate retention tests to make up the 100 questions in each of these tests.

There were 43 questions in the immediate retention test on the first lesson, and 54 questions on the immediate retention test after the second lesson. There were 50 questions in each of the first or immediate retention tests following the third and fourth lessons.

In order to eliminate errors all tests and their directions were carefully checked by several persons in the field of tests and measurements.

*Time of Experiment*

The pre-test and the last two retention tests required not more than 30 minutes to answer. Each lesson was approximately 15 minutes in length and its first or immediate retention test took about 15 minutes of the pupils' time. Therefore, no period of testing, or teaching and testing, was over 30 minutes in length.

The experiment took place during the spring semester of the 1936–37 school year. The administrative staff and faculty of the Lincoln Schools were very cooperative throughout the experiment.

*Subject Matter*

The subject matter chosen for the lessons was selected from a course of study on narcotics, constructed by the writer with the assistance of a committee of junior high school teachers. The committee exercised care in placing the materials on the proper grade levels. Therefore, since the lessons for this experiment were taken directly from the course, the lessons were on the learners' level.

*Lessons*

Because the lecture or straight-talk type of presentation is very common in the class room and over the radio it was chosen as the type of presentation for this experiment. The purpose of the first lesson was to establish some basic concepts concerning narcotics, their advertising, and the qualities to expect in an authority on narcotics. The purpose of the second and third lessons was to present certain factual material concerning alcohol and tobacco. The purpose of the fourth lesson was to present facts concerning narcotics other than alcohol and tobacco. Suggestions by Lumley[5] and Denison[6] concerning the construction and giving of talks to be heard over the loud speaker were followed as they appeared to apply to this experimental situation. In no way, however, were the lessons made to be super-impressive, as many talks are, which one is likely to hear over the radio. The lessons more nearly resembled talks one is likely to hear in the ordinary class room.

*Time Between Tests*

The pupils knew they were to be tested on the lessons and whenever they asked questions about their scores they were encouraged and were promised their scores after the last test.

The pre-test was given from 1 to 3 days before the first lesson. The time between lessons one and two was 1 to 4 days; between lessons two and three, 1 to 3 days; and between lessons three and four, 1 to 5 days. The first retention tests were given immediately following each lesson. The second retention test was given 21 to 28 days after the last lesson. The third and last retention test was given 16 to 21 days after the second retention test. The last retention test was given an average of 44 days after the last lesson.

*Presentation of Lessons*

The lessons were presented in the same way in the ten experimental set-ups. The speaker stood so that he would be about eight inches from the microphone and could talk in a natural tone. No effort was made to be impressive or oratorical. However, an effort was made to accentuate the thought or idea in each sentence. Care was used to pronounce words clearly and distinctly.

*Equipment*

The amplifying unit, loud speaker, microphone, and cable were all used as an amplifying system a year prior to their use in the experiment. The microphone was a high grade carbon microphone which worked very satisfactorily with the other equipment. Care was exercised in the pronunciation of words which carried an "s" or an "sh" sound, since the carbon microphone has a tendency to

[5] F. H. Lumley, "Radio's Rate of Speech," *Educational Research Bulletin*, XII (November, 1933), 244–46.

[6] Merrill Denison, *The Educational Program*. New York: The Radio Institute of the Audible Arts, 1936. Pp. 11.

make such sounds annoying from the loud speaker. The experimenter had experience talking over this particular microphone several times a month for a year prior to this experiment. The reports from the children and the teachers indicated that the loud speaker productions were clear and audible. The microphone cable was shielded in an effort to eliminate interference. The loud speaker was an eight inch dynamic speaker, common at the present time in most of the ordinary home radio sets. The amplifying unit was capable of developing twelve watts, which, of course, is a great deal more than is needed to run one loud speaker.

During the entire experiment the tone control and power rheostat were set so that the quality and volume of the voice heard over the loud speaker were the same for all groups. There was a slight transformer hum audible in any part of the reception rooms. While this was undesirable as far as making the loud speaker presentation perfect is concerned, it was the same for all groups. Pupils reported that the hum was not objectionable and did not interfere with their understanding. Since the schools in Lincoln did not have public address systems it was necessary to have portable equipment for this experiment. It was easily set up and very durable.

*Experimental Technique*

Many difficulties were encountered in attempting to get experimental groups in the various schools which would continue as whole groups throughout the time necessary for the experiment. Chief of these difficulties was absences. Actually 639 pupils participated in the experiment, but the results from only 449 were used, principally because of absence which made the records for some incomplete.

Of these 449 pupils 207 were in Group I and 242 were in Group II. The group rotation technique was selected as best for the experiment. The types of presentation were rotated between the groups so that all the pupils heard half the material by each method, alternation taking place throughout.

The letter "A" refers to the presentation given directly from the speaker and the letter "B" refers to the presentation over the loud speaker. The results of each, A and B, are based on two test responses per pupil. Therefore, where A and B groups are compared, there are twice as many pupil responses as there are actually pupils taking part. For example, 898 subject responses in Group A and 898 subject responses in Group B are compared in Table II. It is well to keep this fact in mind in reading the tables, since the numbers in the tables, except Table I, indicate pupil responses rather than pupils.

*Natural School Conditions*

The conditions under which the experiment took place were natural for the pupils. Regular class groups within the schools, with regular teachers in the rooms, made up the groups for the experiment. These teachers reported that the learning situations were very good.

One person, the writer, gave all the lessons and administered all the tests, thus eliminating the errors that might be introduced by a number of experimenters taking part. Every effort was made to have all experimental set-ups the same.

*The Subjects*

The subjects for the experiment were normal eighth grade pupils from the seven junior high schools located in the various parts of the city. In one school there were three experimental set-ups, and in another school there were two experimental set-ups. However, they are listed as ten different schools rather than set-ups in order not to disclose the identity of the schools.

*Collecting and Computing Data*

The tests were corrected by trained assistants and checked for accuracy. The scoring was done by subtracting the wrongs from the rights and then transposing the answer into a percentage score. All data were collected and recorded on cards. After the data were recorded the information on the card was coded and punched in Hollerith machine cards. A great deal of the computing was done with the use of the Hollerith machine through the International Business Machines Corporation.

## Data Pertaining to the Learning by the Two Types of Presentation

It is the purpose in this section to set forth the findings relative to the learning of the pupils by both types of presentation. Besides showing the Group I and II differences and the way they were combined by the rotation technique to show the Group A and Group B

differences, there are twelve minor studies reported relating to the principal problem.

*Groups I and II*

The means shown in Table I were found by combining the Group I scores from the ten school groups and the Group II scores from the ten school groups. This table makes it possible to study the mean differences between the combined Group I scores and the combined Group II scores. The table is divided into four parts. Each part contains the means on those parts of the tests referring to each of the four lessons. The names of the tests are listed in the first column and the subscripts indicate to which lesson they refer. The actual number of pupils is entered in the third column. Mean differences are found in every case by subtracting the Group I mean from the Group II mean. It is well to keep in mind in reading this table that the group differences have not been equalized. The results, therefore, found in this table are from the groups as they were found in the schools.

The first ratio of the mean difference, divided by the standard deviation of the difference, in the column to the extreme right of the table shows that the difference between Group I and Group II means on the first pre-test is not statistically significant. On the first or immediate retention test Group I is superior, and their superiority is statistically significant, as shown by the corresponding ratio in the last column. Group I heard this lesson directly from the speaker and Group II heard the lesson over the loud speaker. On both of the last retention tests on this lesson the ratios indicate that the group differences are not statistically significant.

In the second part of the table on the pre-test$_2$ the mean difference is statistically in-

TABLE I

A COMPARISON BY LESSONS OF GROUP I AND GROUP II TEST SCORE MEANS BY A AND B PRESENTATIONS

| Test | Group | No. | Mean | Standard Deviation | $\sigma$ m | Diff. | $\sigma$ Diff. | $\frac{\text{Diff.}}{\sigma \text{ Diff.}}$ |
|---|---|---|---|---|---|---|---|---|
| Pre-test$_1$ | *I | 207 | 47.94 | 18.7 | 1.34 | | | |
| | II** | 242 | 47.98 | 13.6 | .90 | —.04 | 1.6 | —.02 |
| Retention Test I$_1$ | I | 207 | 61.66 | 20.0 | 1.41 | | | |
| | II | 242 | 55.54 | 18.0 | 1.2 | 6.12 | 1.85 | 3.31 |
| Retention Test II$_1$ | I | 207 | 63.82 | 17.4 | 1.24 | | | |
| | II | 242 | 64.02 | 17.7 | 1.18 | —.20 | 1.7 | —.12 |
| Retention Test III$_1$ | I | 207 | 63.33 | 16.3 | 1.16 | | | |
| | II | 242 | 61.26 | 22.6 | 1.5 | 2.07 | 1.9 | 1.09 |
| Pre-test$_2$ | I** | 207 | 33.84 | 17.3 | 1.23 | | | |
| | *II | 242 | 35.38 | 17.4 | 1.16 | —1.54 | 1.68 | —.92 |
| Retention Test I$_2$ | I | 207 | 52.15 | 21.1 | 1.5 | | | |
| | II | 242 | 63.66 | 19.4 | 1.29 | —11.51 | 1.96 | —5.87 |
| Retention Test II$_2$ | I | 207 | 41.47 | 16.9 | 1.2 | | | |
| | II | 242 | 49.85 | 20.6 | 1.36 | —8.38 | 1.8 | —4.66 |
| Retention Test III$_2$ | I | 207 | 42.21 | 19.1 | 1.36 | | | |
| | II | 242 | 50.38 | 20.4 | 1.36 | —7.17 | 2. | —3.59 |
| Pre-tests$_3$ | I** | 207 | 33.45 | 16.8 | 1.2 | | | |
| | *II | 242 | 31.85 | 16.2 | 1.07 | 1.60 | 1.6 | 1. |
| Retention Test I$_3$ | I | 207 | 58.72 | 27.0 | 1.9 | | | |
| | II | 242 | 67.07 | 26.0 | 1.7 | —8.35 | 2.55 | —3.27 |
| Retention Test II$_3$ | I | 207 | 43.79 | 17.4 | 1.23 | | | |
| | II | 242 | 46.79 | 12.8 | .85 | —3.00 | 1.5 | —2. |
| Retention Test III$_3$ | I | 207 | 43.68 | 18.7 | 1.33 | | | |
| | II | 242 | 46.67 | 17.9 | 1.19 | —2.99 | 1.78 | —1.68 |
| Pre-test$_4$ | *I | 207 | 37.18 | 19.3 | 1.38 | | | |
| | II** | 242 | 39.33 | 28.7 | 1.91 | —2.25 | 2.35 | —.96 |
| Retention Test I$_4$ | I | 207 | 60.83 | 20.6 | 1.41 | | | |
| | II | 242 | 56.52 | 17.0 | 1.13 | 4.31 | 1.8 | 2.39 |
| Retention Test II$_4$ | I | 207 | 62.55 | 20.1 | 1.42 | | | |
| | II | 242 | 61.81 | 20.9 | 1.39 | .74 | 1.98 | .37 |
| Retention Test III$_4$ | I | 207 | 66.69 | 18.8 | 1.34 | | | |
| | II | 242 | 66.03 | 19.5 | 1.3 | .66 | 1.85 | .36 |

significant, but for the three retention tests Group II, which heard lesson two directly, has significantly higher scores.

The superiority of this method is again demonstrated in these unmatched group scores by the retention test scores following the last two lessons shown in the last two parts of the table.

Since no effort was made to allow for the differences in Group I and Group II in recording and computing these scores, the superiority of the direct type of presentation may not be real, due to the differences in the groups. Likewise, retention test scores are colored by differences on the pre-test scores. In recording the data in this way no allowances have been made for any test or lesson difficulties which may have existed.

In order to get a more accurate, simpler, and a more nearly composite picture of the results from the two types of presentation, Table II is constructed to show the mean of the combined test scores covering the material given by the direct method, compared with the means of the test scores covering the material given by the loud speaker presentation combined. These scores are combined following the rotation technique described in the section on "Procedure." The combination of the scores is indicated in Table I by the asterisks in the column headed, "Group." The single asterisk is placed by the four divisions of the pre-test combined to show in Table II the mean A score on that part of the pre-test covering the material heard later directly from the speaker by both groups. The double asterisk is placed by those pre-test group scores combined to show in Table II the mean B score on that part of the four divisions of the pre-test covering the material heard later over the loud speaker by both groups. The same combination is applied to the retention tests to show the combined means in Table II.

*Groups A and B*

In Table II the differences of the mean test scores on the types of presentation, A and B, are shown. On the pre-test, covering presentation A material, the group had a mean score of 37.74, with a standard deviation of 17.88. The test covering the B type of presentation material had a mean score on the pre-test of 39.04, with a standard deviation of 18.95. The mean differences, where A and B averages are compared, are found by subtracting the mean of B from the mean of A. The mean difference between the two groups on the pre-test is —1.3, which indicates that the difference is in favor of the material given by A, or the direct presentation. Since the mean difference divided by the standard deviation of the difference is —.72, as shown in the last column in the table, there is no significant statistical difference between the groups on this test prior to their hearing the lessons.

On the immediate, or first retention tests, given immediately following each lesson by the A presentation, the group averaged higher than by the B presentation. The ratio resulting from the difference divided by the standard deviation of the differences is 3.78 which shows the difference between the averages to be decidedly significant. It appears that the experimental factor, namely, that of the presence of the speaker is responsible for this impressive difference. On the last two retention tests by the A presentation the group averaged higher than by the B, or loud speaker presentation, but the ratios indicate that their difference is not significant.

It is interesting to note that the marked difference at the end of the first retention test

TABLE II

A COMPARISON OF MEANS ON TESTS COVERING LESSONS GIVEN BY THE A, OR DIRECT TYPE OF PRESENTATION, AND THE B, OR LOUD SPEAKER TYPE OF PRESENTATION

| Test | Group | Mean | Standard Deviation | $\sigma_m$ | Diff. | $\sigma$Diff. | $\frac{\text{Diff.}}{\sigma\text{Diff.}}$ |
|---|---|---|---|---|---|---|---|
| Pre-test | A | 37.74 | 17.88 | 1.24 | —1.30 | 1.80 | —.72 |
| | B | 39.04 | 18.95 | 1.31 | | | |
| Retention Test I | A | 63.47 | 21.51 | 1.45 | 7.71 | 2.04 | 3.78 |
| | B | 55.76 | 20.79 | 1.43 | | | |
| Retention Test II | A | 55.18 | 17.73 | 1.22 | 1.61 | 1.75 | .92 |
| | B | 53.57 | 18.23 | 1.25 | | | |
| Retention Test III | A | 56.12 | 18.37 | 1.26 | 1.63 | 1.80 | .91 |
| | B | 54.49 | 19.99 | 1.37 | | | |

disappears to a large extent by the time of the last retention test. The disappearance of the difference between the two means might be due to a net gain of the low mean or a net loss of the high mean. In this case it is due to a net loss of the high mean, namely, the one for the group by the A presentation. On the first retention test the mean indicates that the pupils by the A presentation knew on an average 63.47 per cent of the material presented, but by the time of the last retention test they had lost 11 per cent of what they knew. On the other hand, the group by the B presentation, the ones hearing the material over the loud speaker, lost only 2 per cent of what they apparently knew at the time of the first retention test. Therefore, while the group hearing the material directly from the speaker in his presence learned more than the group hearing the material over the loud speaker they likewise forgot more; in fact, so much that by the end of the third retention test, which was given 44 days after the first retention test, the difference between the learning of the two groups, as indicated by the means on this test, is not statistically significant.

Assuming that the mean difference in favor of the material covered by the B presentation on the pre-test is an advantage throughout the retention tests, it would increase the difference divided by the standard deviation of the difference ratio on these scores. This would make the superiority of the A presentation even more apparent, but it would not change the ratios on the last two retention tests sufficiently to make the difference statistically significant.

The group by the A presentation made the greatest net learning gain for the period of the experiment. They showed a profit of 32.75 per cent more of the material at the end of the experiment. By the B presentation, the group that heard the material over the loud speaker showed a profit of 28.35 per cent at the end of the experiment. The group by the A presentation showed a profit of 4.40 per cent more than they did by the B presentation.

By both the A and B presentations there is an increase in the mean score on the third retention test over the mean score on the second retention test. This two per cent increase for both groups is probably due to reminiscence. It is not due entirely to reading related material between the two retention tests. This is shown in Table III.

There was no suggestion in the lessons to the pupils that they read related information. However, it was anticipated that some of them might read and thus affect their scores on the last two retention tests. Question 501 appeared in the second retention test and it asked pupils to answer in the affirmative if they had read related material between the lessons and the time of the second retention test. Similarly, question 502 on the third retention test asked pupils to answer in the affirmative if they had read related material in the interval between the two last retention tests.

Table III makes it possible to study the A and B mean differences for the pupils who read between the last two retention tests and the group as a whole. As can be seen by the table, this group averaged higher than the group as a whole, of which they were a part. However, their superiority is not statistically significant. Their superiority is not sufficient to explain the increase in the mean score of the third retention test over the second retention test for the group as a whole. Therefore, this increase for the whole group may be due to reminiscence.

When it is considered that reminiscence might occur in this special group of pupils as well as the group as a whole, some of the amounts of the scores may be due to this phenomenon. The intelligence of the groups was approximately the same. The group as

**TABLE III**

MEAN DIFFERENCES BY THE A AND B PRESENTATIONS FOR PUPILS WHO READ BETWEEN THE LAST RETENTION TESTS AND THE GROUP AS A WHOLE

| Test | Group | No. | Mean | Standard Deviation | $\sigma_m$ | Diff. / $\sigma_{Diff.}$ |
|---|---|---|---|---|---|---|
| Retention Test III | (A—Pupils Who Read) | 176 | 58.40 | 23.50 | 1.77 | 1.05 |
| | (A—Group as a Whole) | 898 | 56.12 | 18.37 | 1.26 | |
| | (B—Pupils Who Read) | 176 | 54.51 | 19.35 | 1.46 | .01 |
| | (B—Group as a Whole) | 898 | 54.49 | 19.99 | 1.37 | |

a whole had an average intelligence quotient of 107; this smaller group, 108. The reading that they did must have supported the material learned directly from the speaker more than the material learned over the loud speaker.

The group of pupils who reported reading between the last two retention tests is smaller than the group who reported reading between the first two retention tests. Therefore, it seems that the question concerning the reading of related information in the first retention test stimulated few pupils to read between the last two retention tests. A comparison of the A and B mean scores for those who read related information in the interval between the first retention test and the second retention test with the group as a whole is shown in Table IV.

In three of the four comparisons of means this special group made higher averages than the group as a whole; however, their superiority is not statistically significant. It is interesting to note that for this special group their B scores are higher than their A scores, which must be due to some factor such as reading. The fact that they average below the group as a whole on that part of the retention test III covering the material presented directly, and that the difference between the averages is almost equal to three times the standard error of the difference, seems to indicate that some factor such as reading has played a part. It is entirely possible that the reading of related information might reduce the scores, as well as raise them, since there is considerable literature easily available in opposition to the facts given in the lessons, as well as supporting such facts. In this regard the reading seemed to increase the standard deviation, indicating a wider variation within this group than the variation in the group as a whole.

*School Marks*

The comparison of mean scores for those pupils making high averages and those making low averages in their school marks is shown in Table V.

The marks in the Lincoln schools are on a seven point scale. A mark of "1" is highest, a mark of "7" is failure. The high mark group includes those who have average marks of either "1's" or "2's", and in the low mark group are those averaging "5's", "6's" and "7's". The junior high school marks, if plotted, would make a skewed curve, as can be seen by the number of pupils representing each level. The upper group contained 111 pupils, the low group 40 pupils, and the middle group 298 pupils.

The pre-test scores, covering the material presented by the loud speaker, are superior for both groups. On all the retention tests the mean scores by the A type of presentation are superior for both groups.

The (difference divided by the standard deviation of the difference) ratios, shown in the column to the extreme right in the table, indicate that the pupils with high marks show a greater difference by the two types of presentation than do those with low marks, with the exception of the last retention test where there is a greater difference in the means for the pupils with low marks, as indicated by their ratio of 1.20. This greater difference of the means for the low group on the third retention test is possibly due to a difference between the groups in reminiscence. For the low mark group there is an increase of 2.76 in the mean score on the third retention test over the mean score of the second retention

TABLE IV

MEAN DIFFERENCES BY THE A AND B PRESENTATIONS FOR PUPILS WHO READ BETWEEN THE FIRST RETENTION TESTS AND THE GROUP AS A WHOLE

| Test | Group | No. | Mean | Standard Deviation | $\sigma_m$ | $\frac{\text{Diff.}}{\sigma\text{Diff.}}$ |
|---|---|---|---|---|---|---|
| Retention Test II | (A—Pupils Who Read) | 214 | 55.43 | 19.50 | 1.33 | .17 |
| | (A—Group as a Whole) | 898 | 55.12 | 17.72 | 1.22 | |
| | (B—Pupils Who Read) | 214 | 56.55 | 21.10 | 1.38 | 1.60 |
| | (B—Group as a Whole) | 898 | 53.57 | 18.23 | 1.25 | |
| Retention Test III | (A—Pupils Who Read) | 214 | 50.85 | 18.00 | 1.23 | 2.99 |
| | (A—Group as a Whole) | 898 | 56.12 | 18.37 | 1.26 | |
| | (B—Pupils Who Read) | 214 | 57.57 | 21.60 | 1.48 | 1.52 |
| | (B—Group as a Whole) | 898 | 54.49 | 19.99 | 1.37 | |

TABLE V

A COMPARISON OF TEST MEANS FOR PUPILS OF HIGH AND LOW SCHOOL MARKS BY A AND B PRESENTATIONS

| Test | Group | Level | No. | Mean | Standard Deviation | σm | Diff. | σDiff. | Diff./σDiff. |
|---|---|---|---|---|---|---|---|---|---|
| Pre-test | A | High | 222 | 41.02 | 17.35 | 1.16 | —2.71 | 1.78 | —1.52 |
| | B | High | 222 | 43.73 | 20.15 | 1.35 | | | |
| | A | Low | 80 | 32.90 | 19.90 | 1.34 | —.90 | 1.90 | —.47 |
| | B | Low | 80 | 33.80 | 20.15 | 1.35 | | | |
| Retention Test I | A | High | 222 | 73.39 | 15.15 | 1.02 | 8.43 | 1.53 | 5.51 |
| | B | High | 222 | 64.96 | 17.00 | 1.14 | | | |
| | A | Low | 80 | 46.41 | 21.05 | 1.41 | 4.23 | 1.97 | 2.15 |
| | B | Low | 80 | 42.18 | 20.35 | 1.37 | | | |
| Retention Test II | A | High | 222 | 62.82 | 17.00 | 1.14 | 1.65 | 1.73 | .95 |
| | B | High | 222 | 61.17 | 19.30 | 1.30 | | | |
| | A | Low | 80 | 41.39 | 23.75 | 1.59 | 1.59 | 2.13 | .75 |
| | B | Low | 80 | 39.80 | 21.10 | 1.42 | | | |
| Retention Test III | A | High | 222 | 63.43 | 17.95 | 1.20 | .76 | 1.80 | .42 |
| | B | High | 222 | 62.67 | 20.00 | 1.34 | | | |
| | A | Low | 80 | 44.15 | 22.65 | 1.52 | 2.56 | 2.14 | 1.20 |
| | B | Low | 80 | 41.59 | 22.55 | 1.51 | | | |

test for the A type of presentation, and only 1.79 points difference between the mean scores of the two retention tests for the B type of presentation.

The net learning gain for the high group is 33 per cent, as shown by the difference between their averages on the pre-test and the last retention test for the A type of presentation. For the low mark group the net gain is 26 per cent, as shown by the difference between the averages of the pre-test and the last retention test for the A type of presentation. According to this interpretation, the high mark group profited more than the low mark group by having the speaker in their presence.

For the presentation over the loud speaker the high mark group showed a profit of 31 per cent, only two per cent less than the gain made by the direct presentation. The low mark group by the B type of presentation showed a gain of 18 per cent, or eight per cent less than by the direct method.

The forgetting calculated from this table corresponds to the findings in the summary Table II concerning forgetting. Of what they knew of the material presented to them directly, the high mark group forgot 14 per cent in the interval between the first retention test and the last retention test, a period of 44 days. Of what they knew of the material presented to them directly, the low mark group forgot only three per cent in the interval between the first retention test and the last retention test, a period of 44 days. For the material learned over the loud speaker the high mark group forgot on the last retention test only four per cent of what they knew at the end of the first retention test.

Marks when grouped in this way are a selective factor in learning by these types of presentation. The pupils with high marks are less affected in their learning by the loud speaker than the low mark group. However, the high mark group forgot a higher percentage of the material learned by the direct method of presentation.

*Averages by Schools*

The averages by schools were studied, which indicated that the schools differed on the amounts they knew before the experiment started, and that the general trends as reported concerning the whole group are present in the separate school groups.

*Intelligence Levels Compared*

Do pupils of a higher intelligence quotient continue to show their superiority in learning from these types of presentation, and how do those of different intelligence levels respond to the experimental factor? Data on the questions are shown in Table VI. The pupils' scores were grouped according to three intelligence quotient levels. The low group included the scores made by those whose intelligence quotients were 94 and below. The

TABLE VI

MEAN TEST SCORE DIFFERENCES BETWEEN DIFFERENT INTELLIGENCE LEVELS FOR A AND B PRESENTATIONS

| Test | Group | I.Q. | No. | Mean | Standard Deviation | σm | Diff. | σDiff. | Diff. / σDiff. |
|---|---|---|---|---|---|---|---|---|---|
| Pre-test | A | Low | 156 | 31.38 | 17.15 | 1.37 | 5.33 | 1.74 | 3.06 |
| | | Med. | 332 | 36.71 | 19.40 | 1.06 | 4.29 | 1.38 | 3.11 |
| | | High | 410 | 41.00 | 17.80 | .88 | 9.62 | 1.63 | 5.90 |
| | B | Low | 156 | 29.74 | 17.35 | 1.39 | 8.41 | 1.74 | 4.83 |
| | | Med. | 332 | 38.16 | 19.15 | 1.05 | 4.71 | 1.37 | 3.44 |
| | | High | 410 | 42.87 | 17.80 | .88 | 13.13 | 1.64 | 8.01 |
| Retention Test I | A | Low | 156 | 47.97 | 19.15 | 1.53 | 11.83 | 1.79 | 6.61 |
| | | Med. | 332 | 59.80 | 16.70 | .92 | 12.53 | 1.19 | 10.53 |
| | | High | 410 | 72.33 | 15.15 | .75 | 24.36 | 1.70 | 14.33 |
| | B | Low | 156 | 40.73 | 18.65 | 1.49 | 11.51 | 1.73 | 6.65 |
| | | Med. | 332 | 52.24 | 16.10 | .88 | 12.09 | 1.20 | 10.08 |
| | | High | 410 | 64.33 | 16.15 | .80 | 23.60 | 1.26 | 18.73 |
| Retention Test II | A | Low | 156 | 42.07 | 19.80 | 1.59 | 10.97 | 1.92 | 5.71 |
| | | Med. | 332 | 53.04 | 19.45 | 1.07 | 8.84 | 1.37 | 6.45 |
| | | High | 410 | 61.88 | 17.30 | .85 | 19.81 | 1.80 | 11.01 |
| | B | Low | 156 | 39.23 | 21.00 | 1.68 | 12.75 | 1.99 | 7.59 |
| | | Med. | 332 | 51.98 | 19.55 | 1.07 | 8.33 | 1.40 | 5.95 |
| | | High | 410 | 60.31 | 18.40 | .91 | 21.08 | 1.91 | 11.04 |
| Retention Test III | A | Low | 156 | 45.07 | 20.40 | 1.63 | 7.91 | 1.95 | 4.06 |
| | | Med. | 332 | 52.98 | 19.75 | 1.08 | 9.89 | 1.39 | 7.12 |
| | | High | 410 | 62.87 | 17.90 | .88 | 17.87 | 1.85 | 9.66 |
| | B | Low | 156 | 39.53 | 21.60 | 1.73 | 13.37 | 2.05 | 6.52 |
| | | Med. | 332 | 52.90 | 20.30 | 1.11 | 8.57 | 1.44 | 5.95 |
| | | High | 410 | 61.47 | 18.55 | .92 | 21.95 | 1.96 | 11.20 |

middle group consisted of all the scores between 95 and 109, and the high intelligence group those with intelligence quotients of 110 and above.

In reading the table for the mean differences, the standard errors, and the (difference divided by the standard deviation of the difference) ratios, the first number in each of these columns is the middle group compared with the low group. The second number is the high group compared with the middle group, and the third number is the high group compared with the low group.

In every case intelligence is a selective factor; that is, those higher in intelligence made higher scores on the tests. The ratios in every case show the differences are statistically significant.

*Intelligence and Presentations*

In Table VII mean scores for the tests covering material taught by each type of presentation for each intelligence level are compared.

On the pre-test the mean of the low group shows that they did not know as much of the material presented to them over the loud speaker as they did of the material presented directly with the speaker in their presence. In the case of the middle group and the upper group the reverse is true. They knew more of the material presented over the loud speaker. However, in no group is the difference between the means statistically significant.

On the first retention test the differences between the mean scores for each level are significant in favor of the direct method of presentation. On the last two retention tests the mean score differences are not significant, as shown by the ratios in the column to the extreme right of the table. They came closer to being significant on the second retention test for the low level group.

The net learning gain on the A type of presentation for the low group, calculated from the pre-test mean to the second retention test mean, is 30 per cent for the middle group. The net gain is 31 per cent, and for the high group the net gain is 35 per cent. The net learning gain for the B type of presentation, calculated in the same way for the low group, is 24 per cent; for the middle group, 26 per cent; and for the high group,

TABLE VII

TEST SCORES COMPARED FOR EACH INTELLIGENCE QUOTIENT LEVEL FOR A AND B PRESENTATIONS

| Test | Group | I.Q. | No. | Mean | Standard Deviation | σm | Diff. | σDiff. | Diff./σDiff. |
|---|---|---|---|---|---|---|---|---|---|
| Pre-test | A | Low | 156 | 31.38 | 17.15 | 1.37 | 1.64 | 1.95 | .84 |
| | B | Low | 156 | 29.74 | 17.35 | 1.39 | | | |
| | A | Med. | 332 | 36.71 | 19.40 | 1.06 | —1.45 | 1.49 | —.97 |
| | B | Med. | 332 | 38.16 | 19.15 | 1.05 | | | |
| | A | High | 410 | 41.00 | 17.80 | .88 | —1.87 | 1.24 | —1.50 |
| | B | High | 410 | 42.87 | 17.80 | .88 | | | |
| Retention Test I | A | Low | 156 | 47.97 | 19.15 | 1.53 | 7.24 | 2.14 | 3.40 |
| | B | Low | 156 | 40.73 | 18.65 | 1.49 | | | |
| | A | Med. | 332 | 59.80 | 16.70 | .92 | 7.56 | 1.27 | 5.95 |
| | B | Med. | 332 | 52.24 | 16.10 | .88 | | | |
| | A | High | 410 | 72.33 | 15.15 | .75 | 8.00 | 1.10 | 7.27 |
| | B | High | 410 | 64.33 | 16.15 | .80 | | | |
| Retention Test II | A | Low | 156 | 42.07 | 19.80 | 1.59 | 2.84 | 2.31 | 1.23 |
| | B | Low | 156 | 39.23 | 21.00 | 1.68 | | | |
| | A | Med. | 332 | 53.04 | 19.45 | 1.07 | 1.06 | 1.51 | .70 |
| | B | Med. | 332 | 51.98 | 19.55 | 1.07 | | | |
| | A | High | 410 | 61.88 | 17.30 | .85 | 1.57 | 1.25 | 1.26 |
| | B | High | 410 | 60.31 | 18.40 | .91 | | | |
| Retention Test III | A | Low | 156 | 45.07 | 20.40 | 1.63 | 5.54 | 2.38 | 2.33 |
| | B | Low | 156 | 39.53 | 21.60 | 1.73 | | | |
| | A | Med. | 332 | 52.98 | 19.75 | 1.08 | .08 | 1.55 | .05 |
| | B | Med. | 332 | 52.90 | 20.30 | 1.11 | | | |
| | A | High | 410 | 62.87 | 17.90 | .88 | 1.40 | 1.27 | 1.10 |
| | B | High | 410 | 61.47 | 18.55 | .92 | | | |

35 per cent. The middle and lower groups, therefore, profited more by the speaker method than they did by the loud speaker method. The high group profited about the same for each method.

In terms of loss, as shown by the difference between the first retention test and the last retention test, by the A method of presentation the low group lost six per cent, the middle group lost 31 per cent, and the high group lost 13 per cent. The loss for the B type of presentation, calculated the same way, for the low group is three per cent, for the middle group one per cent, and for the high group four per cent.

In five out of the six cases the third retention test scores are higher than the second retention test scores, again showing that reminiscence has occurred.

*Sex*

Table VIII shows the mean scores of the boys compared with the mean scores of the girls. It is indicated in every case that the test averages for the boys are higher than for the girls. This suggests that possibly the topic of narcotics has been more of a boy's problem.

The differences in their averages are significant on the immediate or first retention test only. However, on the pre-test, covering information taught by the direct method, the boys' superiority reduced to chances is 98 in 100, and by material taught by the B type, the boys are superior 99 chances in 100.

On the second retention test the boys are superior on the material taught by the direct method 99 chances in 100, and by the loud speaker method 93 chances in 100.

On the last retention test, covering material taught by the A type of presentation, the boys are superior by 98 chances in 100, but on the test covering the B type of presentation the boys are superior only 65 chances in 100.

The odds, then, are definitely in favor of the boys in this experiment, if one does not allow for their superiority in the pre-test in the remainder of the tests. If their pre-test superiority is an advantage throughout the retention tests, then the net learning gain or profit, as shown between the pre-test and

TABLE VIII

MEAN TEST SCORE DIFFERENCES BY SEX FOR A AND B PRESENTATION

| Test | Group | Sex | No. | Mean | Standard Deviation | σm | Diff. | σDiff. | Diff./σDiff. |
|---|---|---|---|---|---|---|---|---|---|
| Pre-test | A | Boys | 454 | 38.82 | 18.6 | .87 | 2.52 | 1.26 | 2.00 |
| | | Girls | 444 | 36.30 | 19.15 | .91 | | | |
| | B | Boys | 454 | 40.82 | 19.15 | .90 | 3.60 | 1.28 | 2.81 |
| | | Girls | 444 | 37.22 | 19.15 | .91 | | | |
| Retention Test I | A | Boys | 454 | 65.80 | 18.35 | .86 | 4.72 | 1.24 | 3.81 |
| | | Girls | 444 | 61.08 | 18.75 | .89 | | | |
| | B | Boys | 454 | 58.43 | 18.55 | .87 | 5.30 | 1.25 | 4.24 |
| | | Girls | 444 | 53.03 | 18.65 | .89 | | | |
| Retention Test II | A | Boys | 454 | 56.75 | 21.65 | 1.02 | 3.19 | 1.35 | 2.36 |
| | | Girls | 444 | 53.56 | 18.85 | .89 | | | |
| | B | Boys | 454 | 54.34 | 22.20 | 1.04 | 2.04 | 1.42 | 1.44 |
| | | Girls | 444 | 52.30 | 20.30 | .96 | | | |
| Retention Test III | A | Boys | 454 | 57.45 | 20.70 | .97 | 2.73 | 1.34 | 2.04 |
| | | Girls | 444 | 54.72 | 19.40 | .93 | | | |
| | B | Boys | 454 | 54.77 | 21.70 | 1.02 | .57 | 1.42 | .40 |
| | | Girls | 444 | 54.20 | 20.95 | .99 | | | |

third retention test, is in favor of the girls. Making allowances for the advantage of the boys on the pre-test, the girls retained more than the boys on the material learned over the loud speaker, as shown by the last two retention test scores by the B method. The mean intelligence quotient for the boys and the girls was the same.

*Occupational Levels*

The test averages of children of parents, usually the father, on high and low occupational levels are shown in Table IX.

The high occupational level includes classifications in the professional group only. The low occupational level includes those occupations in the unskilled and semi-skilled groups. The classification was made from the occupation of parents listed on the children's school census sheets according to the occupational division set up by Arthur J. Jones.[7] The averages on the low level are possibly more reliable, since this level includes almost twice as many subjects as in the high group. Since

[7] Arthur J. Jones, *Principles of Guidance*, p. 314. New York: McGraw-Hill Book Co., 1930.

TABLE IX

MEAN TEST SCORE DIFFERENCES FOR CHILDREN OF PARENTS ON HIGH AND LOW OCCUPATIONAL LEVELS BY A AND B PRESENTATIONS

| Test | Group | Level | No. | Mean | Standard Deviation | σm | Diff. | σDiff. | Diff./σDiff. |
|---|---|---|---|---|---|---|---|---|---|
| Pre-test | A | High | 72 | 40.15 | 19.35 | 2.28 | —6.00 | 3.20 | —1.88 |
| | B | High | 72 | 46.15 | 19.00 | 2.24 | | | |
| | A | Low | 184 | 36.28 | 19.00 | 1.40 | 2.05 | 1.76 | 1.16 |
| | B | Low | 184 | 34.23 | 14.40 | 1.06 | | | |
| Retention Test I | A | High | 72 | 69.84 | 17.05 | 2.01 | 11.18 | 3.14 | 3.58 |
| | B | High | 72 | 58.66 | 20.44 | 2.41 | | | |
| | A | Low | 184 | 59.11 | 19.95 | 1.47 | 8.54 | 1.97 | 4.34 |
| | B | Low | 184 | 50.57 | 17.75 | 1.31 | | | |
| Retention Test II | A | High | 72 | 61.79 | 18.45 | 2.17 | 3.93 | 3.12 | 1.26 |
| | B | High | 72 | 57.86 | 19.00 | 2.24 | | | |
| | A | Low | 184 | 52.97 | 21.35 | 1.57 | 5.28 | 2.17 | 2.43 |
| | B | Low | 184 | 47.69 | 20.40 | 1.50 | | | |
| Retention Test III | A | High | 72 | 63.26 | 18.35 | 2.16 | 3.40 | 3.08 | 1.10 |
| | B | High | 72 | 59.86 | 18.65 | 2.20 | | | |
| | A | Low | 184 | 55.07 | 20.40 | 1.50 | 7.79 | 2.20 | 3.54 |
| | B | Low | 184 | 47.28 | 21.85 | 1.61 | | | |

both these smaller groups were in the whole group, they were subjected to the same conditions in the experiment and therefore their scores can be compared.

There was less difference in the results of the two methods of presentation for those children from the high occupational level, as shown by the (difference divided by the standard deviation of the difference) ratios. This is to be expected from our study of intelligence and test scores, since the intelligence quotient average for the high occupational group is 113 and for the low occupational group is 101. Even allowing for pre-test advantages, the average scores in every test were higher for the children from families of the higher occupational level.

The percentage of loss from the first retention test to the last retention test is less in this experiment than any of the others. By the direct type of presentation the high occupational group lost nine per cent; the low occupational group lost seven per cent. By the loud speaker type of presentation the high occupational group actually gained two per cent, and the low occupational group lost six per cent. The gain on the part of the high occupational group possibly is due to reminiscence. However, they show a loss on the average in that part of the retention test which covered the material given them by the B or loud speaker type of presentation.

The third retention test scores in three out of four cases are higher than the corresponding second retention test scores. This increase may be due to reminiscence.

The total net learning gain for the high group from the pre-test average to the final test average by the A method is 37 per cent; by the B method, 23 per cent. For the low occupational group the net learning gain as shown between the averages for the A method is 34 per cent; for the B method, is 27 per cent. The low occupational group then, profited more in terms of the percentage learned from the loud speaker presentation than did the high occupational group.

TABLE X

SCHOOL GRADE AVERAGES FOR CHILDREN OF PARENTS IN HIGH AND LOW OCCUPATIONAL LEVELS

| | School Marks | | | | | | |
|---|---|---|---|---|---|---|---|
| Occupational Level | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Children from Professions | 3 | 19 | 9 | 4 | 1 | | |
| Children from Low Occupations | 1 | 9 | 33 | 35 | 9 | 4 | 1 |

Children of parents in professional work averaged two points higher on a seven point scale than children of the lower groups. The relationship of the scores, then, for this group is consistent with the group concerning school marks.

On the high occupational level only one pupil came from a home without a radio; since there were 36 pupils in the group, this pupil represents about three per cent of the group. In the low occupational group nine pupils reported no radio in the home. Five of this group did not report. There were 92 pupils in this group, which means that about ten per cent reported no radio in their homes.

Since the radio group has a lower average on tests over material learned from the loud speaker, the report on the occupational levels is consistent.

About 65 per cent of the children from the high occupational level read material, and about the same percentage of the children

TABLE XI

MEAN TEST SCORE DIFFERENCES FOR CHILDREN OF FOREIGN BORN PARENTS BY A AND B PRESENTATIONS

| Test | Group | No. | Mean | Standard Deviation | $\sigma m$ | Diff. | $\sigma$Diff. | Diff. / $\sigma$Diff. |
|---|---|---|---|---|---|---|---|---|
| Pre-test | A | 86 | 33.95 | 19.50 | 2.10 | .56 | 2.89 | .19 |
| | B | 86 | 33.39 | 18.45 | 1.99 | | | |
| Retention Test I | A | 86 | 55.69 | 20.20 | 2.18 | 9.80 | 3.06 | 3.20 |
| | B | 86 | 45.89 | 19.80 | 2.14 | | | |
| Retention Test II | A | 86 | 48.23 | 22.75 | 2.45 | —.28 | 3.36 | —.08 |
| | B | 86 | 48.51 | 21.35 | 2.30 | | | |
| Retention Test III | A | 86 | 50.91 | 20.85 | 2.24 | 2.68 | 3.44 | .78 |
| | B | 86 | 48.23 | 24.15 | 2.61 | | | |

from the low occupational level read. If the reading affected one group, one would expect it to affect the other group likewise.

It is interesting, however, that more than one half of this group read related material, even though no suggestion to do so was made during the experiment.

None of the foreign born parents is reported in the high occupational or professional group, while 17 per cent of the foreign born are reported in the low occupational level group.

*Foreign Born Parents*

From the school census sheets the birth places of parents were tabulated for those pupils taking part in the experiment. For this study only the pupils were taken whose parents were both reported as being born in some country other than the United States. Most of these parents were born in Russia. A comparison of their means can be found in Table XI.

On the pre-test the mean scores on the material covering each type of presentation are practically the same, at least their (difference divided by the standard deviation of the difference) ratio shows that there is no significant difference. The first retention test averages show a decided advantage in favor of the A type of presentation; reduced to chances this type of presentation is better by 99 chances in 100. On the third and last retention test the means show that the pupils remembered more of what they had learned in the speaker's presence, but they lost some of what they had learned as is indicated on the first retention test. The material they learned over the loud speaker shows a definite gain in the last two retention tests over the first retention test. In one of the two cases the pupils had a higher mean score on the third retention test than on the second. The most obvious explanation for this is reminiscence.

In every test the mean scores of these pupils are lower than the mean scores of the group as a whole, as shown in Table II. These pupils have an intelligence quotient average of 99, which is below the average of the group as a whole. Therefore, the scores reported in this table are consistent with the intelligence study. They also averaged a "4" in their marks on junior high school work. This mark is a low average on a seven point scale where "1" is perfect and "7" is failure. Therefore, the scores of these pupils are consistent with the scores in the section concerning school marks.

These pupils show a net learning gain of 33 per cent from the time of the pre-test to the last retention test on the material they learned by the direct method. This is one per cent more than the gain of the group as a whole, which was 32 per cent. Of the material taught them over the loud speaker, they show a net learning gain of 32 per cent between the pre-test and the last retention test. This is more than the group as a whole gained, which was 28 per cent. This extra gain might be due in part to the novelty of learning over the loud speaker for some of these children, since thirteen per cent of the group reported coming from homes without radios.

*No Radios in the Homes*

The group of pupils reported in Table XII, as not having radios in their homes, averaged below the group as a whole on all scores except the pre-test score covering the material taught by the direct method. Therefore, their averages on succeeding tests may be higher for this one method because of their knowledge prior to the experiment. Not allowing for the pre-test advantage, and reducing the ratios to chances in 100, the superiority in favor of the direct method over the indirect method is 98 chances in 100 on the first retention test, while on the second retention test the direct method is superior by 84 chances in 100.

Due to some additional loss of the material learned by the indirect method and a gain on the material learned by the direct method on the third retention test, the latter method is superior by 93 chances in 100.

On the retention tests they averaged higher on the material learned directly. However, their averages indicate that they forgot more of the material learned directly, which reduces the difference between the averages on the material learned from the two methods.

Had the (difference divided by the standard deviation of the difference) ratio on the pre-test been reduced to zero by reducing their advantage on the material presented by the A type of presentation, undoubtedly the differences of the means on the following tests, shown by the ratios, would have been changed. Assuming that the advantage in favor of the direct lesson material on the pre-test played a constant part on the follow-

ing test scores, the indirect method would be superior on the second retention test by 65 chances in 100. The direct method would be superior on the first retention test by 74 chances in 100, and superior on the second retention test by only 53 chances in 100, thus making all of the mean differences statistically insignificant.

The net learning gain on the material learned by the A type of presentation is 23 per cent, which is less than for the group as a whole by nine per cent. The net learning gain on the material learned from the loud speaker presentation is 24 per cent. This is four per cent less than for the group as a whole. In proportion, then, to the whole group these pupils really turned in a better performance on the tests covering the material they heard over the loud speaker. It is to be expected that the mean scores of this group would be below the average of the group as a whole, since ten per cent of this group came from the low intelligence quotient group and sixty per cent came from the middle intelligence quotient group. Only 43 per cent of the group as a whole were in the lower and middle intelligence quotient groups. Also, more of this group came from the low occupational group.

## Summary and Conclusions

This study was concerned with the learning and retention of factual material presented over the loud speaker to four hundred and forty-nine normal eighth grade pupils compared with their learning and retention of factual material presented directly by the speaker himself. The content of the factual material concerned narcotics.

After a pre-test was given, four lessons were presented to the pupils on which they were tested immediately after each lesson and at two other dates, later by twenty-four and twenty days, respectively.

In presenting each lesson in each of the ten schools there were two groups of pupils. One group heard a lesson directly from the speaker with the speaker present, and the other group listened at the same time in another room over the loud speaker. The four presentations were rotated between the groups so that each group heard two talks over the loud speaker and two directly from the speaker. The results were combined according to the types of presentation. The results of the two types of presentation with the group as a whole were studied, and in connection with several other factors common to the whole group or parts of the group. Natural school room conditions prevailed throughout the experiment.

The pupils knew a similar amount of the material presented to them by the two types, as shown by the mean scores on the pre-test. On the immediate retention tests the mean differences between the amounts learned by the two types of presentation are significant in favor of the direct type of presentation, that is, having the material presented to the pupils by the speaker being present. However, on the last two retention tests the mean differences indicate that the two types of presentation are almost equally effective. At least, the (difference divided by the standard deviation of the difference) ratios indicate that the differences are not significant.

The pupils made the greatest net learning gain on the material presented to them directly. While there was forgetting of

**Table XII**

**Mean Test Score Differences for Pupils From Homes Without Radios by A and B Presentations**

| Test | Group | No. | Mean | Standard Deviation | $\sigma_m$ | Diff. | $\sigma_{Diff.}$ | $\frac{Diff.}{\sigma_{Diff.}}$ |
|---|---|---|---|---|---|---|---|---|
| Pre-test | A | 60 | 40.73 | 18.50 | 2.40 | 4.73 | 3.30 | 1.43 |
| | B | 60 | 36.00 | 17.15 | 2.08 | | | |
| Retention Test I | A | 60 | 58.97 | 17.45 | 2.25 | 6.55 | 3.13 | 2.09 |
| | B | 60 | 52.42 | 16.85 | 2.17 | | | |
| Retention Test II | A | 60 | 52.85 | 22.00 | 2.84 | 3.72 | 3.62 | 1.02 |
| | B | 60 | 49.13 | 17.35 | 2.24 | | | |
| Retention Test III | A | 60 | 53.00 | 20.75 | 2.67 | 5.93 | 3.92 | 1.51 |
| | B | 60 | 47.07 | 22.25 | 2.87 | | | |

the amounts learned by both types of presentation, a higher percentage of the material learned over the loud speaker was remembered.

In almost every case, apparently due to reminiscence, the last retention test scores were higher than the next to the last retention test scores.

Minor studies connected with this problem are as follows:

1. Even though no suggestion was made to the pupils that they read related material, some reported that they did read in the intervals between the retention tests. Their reading had a tendency to raise their retention test scores above the average of the group, but not sufficiently to make their scores significantly higher than the group as a whole, and not a sufficient amount to explain away reminiscence.

2. Pupils with high average marks in their regular junior high school work learned more by both types of presentation than those with the low average school marks. The results from this group demonstrate again that while pupils remember slightly more by having the speaker present, they forget a smaller percentage of what they learned over the loud speaker.

3. Intelligence scores are a selective factor in learning by both types of presentation. Pupils of high intelligence learn more by both methods than the middle and lower intelligence groups. The net learning gain for those of higher intelligence is about the same for both types of presentation. The middle and lower groups retained more by having the speaker present.

4. The test scores for the boys are higher than for the girls by each type of presentation. The average intelligence for the boys as a group was about the same as for the girls as a group. The pre-test mean scores indicate that the boys knew more at the beginning of the experiment. There was less difference between the two groups as the tests continued. Making allowance for the advantage the boys had at the beginning of the experiment, the girls' scores are about the same as the boys'. The girls showed some superiority in learning by the loud speaker method of presentation.

5. The mean test scores of the group of pupils whose parents, mainly the father, were classed in the professional occupations were compared with the mean scores of those whose parents were from the unskilled and semi-skilled occupations. The pupils in the high occupational group learned more by the direct method than the pupils from the low occupational group. Those in the low occupational group profited more from the loud speaker type of presentation. It is interesting to note that the high occupational group averaged higher in intelligence, and also averaged two points higher in their school marks on the seven point scale. About the same percentage of each of these groups read related material between the retention tests. About ten per cent of the low occupational group reported having no radios in their home. Learning from the loud speaker, therefore, for that group may have been somewhat of a novelty.

6. The learning, by the two types of presentation, of pupils whose parents were born outside of the United States was studied and compared with the group as a whole. Their test scores were lower than the group as a whole, and their net learning gain for the time of the experiment was more than that for the group as a whole by both methods. This might be due in part to the novelty of the loud speaker presentation, because thirteen per cent of this group reported not having radios in their homes. On the whole, this group profited slightly more by having the speaker present, but the differences are statistically insignificant. It is of interest to know that this group was below the average of the group as a whole in intelligence.

7. The pupils who reported that they did not have radios in their homes profited slightly more by the loud speaker presentation for the time of the experiment. On the tests they averaged below the whole group by both methods, which is expected, since they averaged below the group as a whole in intelligence. In school marks they averaged lower than the group as a whole, and a higher percentage came from the low occupational group. In general, it seems that these pupils were not handicapped by learning over the loud speaker.

# RETROACTIVE INHIBITION AS A FUNCTION OF AGE, INTELLIGENCE, AND THE DURATION OF THE INTERPOLATED ACTIVITY[1]

SISTER M. FLORENCE LOUISE LAHEY
*Marygrove College*
*Detroit*

The vast amount of recent experimental research on retroactive inhibition[2] has proved that degree of retention is largely a function of the activities intervening between learning and recall. The disuse hypothesis, therefore, has been discarded as a complete explanation of forgetting, and in its place, has been substituted a theory based upon the inhibitory influence that the learning of an interpolated activity has upon the retention of previously acquired material. Experimental investigations have shown that the amount of inhibition is subject to many conditions involved in and surrounding the learning situation. Previous studies of the phenomenon have been concerned chiefly with such factors as the similarity of the original and interpolated learning, the temporal position of the interpolation, the amount and degree of learning of the original material, the amount and degree of learning of the interpolated activity, the kind of material used for original learning, attitudes and mental sets, emotional disturbances before or during the interpolated study, and method of measurement.

The present study[3] was one of a number of experimental investigations on retroactive inhibition conducted in educational psychology at the Catholic University of America. A serial list of twenty-five verbs, taken from the first twenty-five hundred words of the Thorndike Word List, was studied for five minutes by 3,434 children from the third through the tenth grades of several Detroit schools. By providing an age range of from eight through sixteen years, this procedure rendered possible comparisons of nine different age-levels. Immediately after a four minute test of the original material, the experimental groups devoted varying amounts of time to the study of an interpolated list of fifty nouns. The periods of interpolation employed were four, six, eight, ten, twelve, and fourteen minutes; each study period was followed by a three minute test of the interpolated series. In every condition of the experiment a fixed interval of seventeen minutes was allowed between the end of the initial test and the recall of the original list of verbs. All groups devoted the interval between the end of the designated period of interpolated study and this recall to the group singing of songs well known and apparently enjoyed by all. This activity was chosen because it is one that approximates a group "rest" interval. For the various experimental conditions the duration of this "rest" activity varied from zero to ten minutes. To provide a common basis for the comparison of results, a control group was used. For this group there was no interpolated study; the entire seventeen minute interval between the initial test and the first recall was devoted to the same type of group singing that was employed with the experimental groups. In order to check the persistence of the retroactive effect, a second recall was requested of all groups at the end of twenty-four hours. In the interpretation of the results derived from this procedure, three problems were placed for consideration:

1. Retroactive inhibition as a function of the length of time devoted to interpolated study.
2. The importance of age as a factor in retroaction.
3. The relationship between degree of mental ability and susceptibility to the detrimental influence of interpolated study.

All results given in this summary of the experimental procedure described are based upon data obtained from groups equated on

[1] From the Department of Education, Catholic University, Washington, D. C.

[2] For a comprehensive review of studies of retroactive inhibition see S. H. Britt, "Retroactive Inhibition: A Review of the Literature," *Psychological Bulletin*, XXXII (1935), 381–440.

[3] Sr. M. Florence Louise Lahey, *Retroactive Inhibition as a Function of Age, Intelligence, and the Duration of the Interpolated Activity*. Educational Research Monographs, Vol. X, No. 2. Washington: Catholic University of America, 1937. Pp. 93.

TABLE I

PERCENTAGES OF RETENTION AND OF RETROACTIVE INHIBITION
SEVENTEEN-MINUTE INTERVAL

| | 8 yrs. | 9 yrs. | 10 yrs. | 11 yrs. | 12 yrs. | 13 yrs. | 14 yrs. | 15 yrs. | 16 yrs. |
|---|---|---|---|---|---|---|---|---|---|
| 4-Min. Interp. | | | | | | | | | |
| Control | 78.8 | 86.8 | 94.7 | 90.4 | 89.8 | 90.1 | 89.6 | 91.2 | 84.9 |
| Exper. | 56.3 | 56.4 | 59.6 | 56.4 | 56.0 | 57.7 | 59.2 | 65.6 | 78.7 |
| R. I. | 28.6 | 35.0 | 37.1 | 37.6 | 37.6 | 36.0 | 33.9 | 28.1 | 7.3 |
| 6-Min. Interp. | | | | | | | | | |
| Control | | 90.6 | 97.6 | 92.3 | 90.0 | 89.4 | 89.3 | 89.9 | 85.6 |
| Exper. | | 55.4 | 48.0 | 47.6 | 58.7 | 59.5 | 62.7 | 65.0 | 65.1 |
| R. I. | | 38.9 | 50.8 | 48.4 | 34.8 | 33.4 | 29.8 | 27.7 | 23.9 |
| 8-Min. Interp. | | | | | | | | | |
| Control | 89.7 | 90.3 | 98.9 | 88.4 | 91.4 | 89.9 | 89.9 | 91.5 | 85.8 |
| Exper. | 50.4 | 54.3 | 48.8 | 54.3 | 51.3 | 57.9 | 58.4 | 62.2 | 70.1 |
| R. I. | 43.8 | 39.9 | 50.7 | 38.6 | 43.9 | 35.6 | 35.0 | 32.0 | 18.3 |
| 10-Min. Interp. | | | | | | | | | |
| Control | 83.1 | 88.3 | 94.7 | 89.0 | 92.1 | 89.9 | 90.4 | 86.8 | 85.8 |
| Exper. | 44.2 | 47.9 | 47.6 | 51.4 | 48.3 | 50.6 | 48.1 | 59.4 | 55.7 |
| R. I. | 46.8 | 45.8 | 49.7 | 42.2 | 47.6 | 43.7 | 46.8 | 31.6 | 35.1 |
| 12-Min. Interp. | | | | | | | | | |
| Control | 85.0 | 90.1 | 94.3 | 90.2 | 90.6 | 89.4 | 90.7 | 87.8 | 84.9 |
| Exper. | 41.5 | 52.2 | 47.7 | 44.9 | 48.3 | 53.8 | 58.3 | 57.5 | 63.1 |
| R. I. | 51.2 | 42.1 | 49.4 | 50.2 | 46.7 | 39.8 | 35.7 | 34.5 | 25.7 |
| 14-Min. Interp. | | | | | | | | | |
| Control | 91.6 | 87.0 | 98.3 | 91.0 | 90.6 | 90.0 | 90.5 | 89.0 | ..... |
| Exper. | 41.2 | 49.6 | 41.7 | 40.1 | 50.2 | 47.7 | 51.8 | 56.9 | |
| R. I. | 55.0 | 43.0 | 57.6 | 55.9 | 44.6 | 47.0 | 42.8 | 36.1 | |

the basis of the mean score on the initial test, measures of variability being ignored. The percentage of retroactive inhibition was obtained from the formula:

$$\frac{\%\ \text{retention}_{\text{control}} - \%\ \text{retention}_{\text{experimental}}}{\%\ \text{retention}_{\text{control}}}$$

As shown in Tables I and II, there is a consistent tendency, under the conditions of this experiment, for the amount of retention to decrease and for the percentage of retroactive inhibition to increase with the augmentation of the time devoted to the study of an interpolated list twice as long as the original series.

Though not so reliable a measure of inhibition as that using the inhibitory effect manifested by individual groups, the method of averaging the amounts of retroaction evidenced by all age-groups at each of the interpolation periods offers a convenient summary of the varying degrees of interference produced by the different amounts of interpolated study. The averages for the various periods are as indicated in Table III.

These data indicate the gradual and regular increase in the percentage of the retroactive influence from the four to the fourteen-minute interpolation at the seventeen-minute recall and from the four to the ten, twelve, and fourteen-minute periods at the twenty-four hour recall. According to every method of comparison, the results indicate that the greatest single amount of inhibition is produced by the first four minutes of interpolated study. As interpolation time extends beyond four minutes, however, there is a gradual increase in the inhibitory effect, though relatively small from one interpolation period to the next. When the ten minute interval is reached, there is a decided increase in the retroactive effect. While evident in the ratios, the increment is not so

TABLE II

PERCENTAGES OF RETENTION AND OF RETROACTIVE INHIBITION
TWENTY-FOUR HOUR RECALL

| | 8 yrs. | 9 yrs. | 10 yrs. | 11 yrs. | 12 yrs. | 13 yrs. | 14 yrs. | 15 yrs. | 16 yrs. |
|---|---|---|---|---|---|---|---|---|---|
| 4-Min. Interp. | | | | | | | | | |
| Control | 77.2 | 73.6 | 76.1 | 81.4 | 76.8 | 80.9 | 80.9 | 85.1 | 73.3 |
| Exper. | 44.5 | 43.8 | 43.6 | 44.1 | 48.0 | 50.5 | 56.4 | 59.2 | 58.3 |
| R. I. | 42.4 | 40.5 | 42.7 | 45.8 | 37.5 | 37.6 | 30.3 | 30.4 | 20.5 |
| 6-Min. Interp. | | | | | | | | | |
| Control | | 76.9 | 80.0 | 82.7 | 76.8 | 71.5 | 78.7 | 82.8 | 76.1 |
| Exper. | | 47.2 | 41.8 | 43.6 | 47.3 | 48.4 | 51.5 | 52.3 | 54.6 |
| R. I. | | 38.6 | 47.8 | 47.3 | 38.4 | 32.3 | 34.6 | 36.8 | 28.3 |
| 8-Min. Interp. | | | | | | | | | |
| Control | 84.8 | 76.9 | 81.8 | 79.7 | 78.8 | 80.3 | 79.5 | 85.0 | 76.3 |
| Exper. | 33.3 | 48.7 | 44.6 | 52.8 | 46.8 | 48.3 | 49.8 | 54.5 | 55.1 |
| R. I. | 60.7 | 36.7 | 45.5 | 33.8 | 40.6 | 39.9 | 37.4 | 35.9 | 27.8 |
| 10-Min. Interp. | | | | | | | | | |
| Control | 80.8 | 73.2 | 76.1 | 80.8 | 81.5 | 80.2 | 82.1 | 80.7 | 68.4 |
| Exper. | 26.0 | 36.3 | 38.8 | 39.5 | 46.9 | 47.2 | 43.9 | 49.1 | 50.9 |
| R. I. | 67.8 | 50.4 | 49.0 | 50.6 | 42.5 | 41.1 | 46.5 | 39.2 | 25.6 |
| 12-Min. Interp. | | | | | | | | | |
| Control | 82.1 | 77.8 | 74.4 | 81.6 | 79.1 | 79.1 | 81.7 | 81.5 | 73.3 |
| Exper. | 30.9 | 33.1 | 39.5 | 35.9 | 48.3 | 48.0 | 53.4 | 55.3 | 44.6 |
| R. I. | 62.4 | 57.5 | 46.9 | 56.0 | 38.9 | 39.3 | 34.6 | 32.1 | 39.2 |
| 14-Min. Interp. | | | | | | | | | |
| Control | 86.4 | 75.1 | 80.1 | 81.9 | 79.1 | 80.4 | 81.7 | 81.8 | |
| Exper. | 35.1 | 40.7 | 38.8 | 39.8 | 44.9 | 49.1 | 50.2 | 50.5 | |
| R. I. | 59.4 | 45.8 | 51.6 | 51.4 | 43.2 | 38.9 | 38.6 | 38.3 | |

pronounced from the ten through the fourteen minute intervals, especially in the delayed recall. The explanation of this apparent plateau in the detrimental influence after ten minutes of interpolated study is probably due to the relation between intelligence and retroactive inhibition as shown in the latter part of this summary.

Tables I and II contain the data used in the preliminary study of the influence of age upon retroaction. On first analysis, these data seemed to indicate that the disturbing

TABLE III

AVERAGE PERCENTAGES OF RETROACTIVE INHIBITION (ALL AGES COMBINED)

| | 17-Minute Recall | | 24-Hour Recall | |
|---|---|---|---|---|
| | % R. I. | Number | % R. I. | Number |
| 4-Minute Interpolation | 31.2 | 395 | 36.4 | 395 |
| Control | | 713 | | 713 |
| 6-Minute Interpolation | 36.0 | 394 | 38.0 | 394 |
| Control | | 624 | | 624 |
| 8-Minute Interpolation | 37.5 | 450 | 39.8 | 450 |
| Control | | 663 | | 663 |
| 10-Minute Interpolation | 43.3 | 497 | 45.9 | 497 |
| Control | | 668 | | 668 |
| 12-Minute Interpolation | 41.7 | 436 | 45.2 | 436 |
| Control | | 654 | | 654 |
| 14-Minute Interpolation | 47.8 | 492 | 45.9 | 492 |
| Control | | 667 | | 667 |

effect of interpolated material decreases with increasing age. In the study of the influence of intelligence upon the retroactive effect, however, the results proved that it is not age alone, but age in relation to intelligence that conditions the degree of retroaction in children.

To determine the relation between intelligence and susceptibility to retroaction, the subjects were classified as bright, average, or dull on the basis of IQ's obtained from the administration of the Otis Group Intelligence Test. Because of the limited number of cases in the different age and interpolation groups, it was decided to depart from the usual classification and to use 94 as the maximum IQ for the dull groups and to designate all attaining an IQ above 105 as bright. The average groups were thus included in the range 95–105 inclusively. Though such coarse grouping would be inadvisable for individuals, it was probably sufficiently reliable for the purposes of this study.

To provide for a representative number of cases in the various age-groups, adjoining age levels were combined. This procedure gave combinations of the eight and nine, ten and eleven, twelve and thirteen, and fourteen and fifteen year levels. This combination into four age levels rendered possible at each recall seventy-two comparisons with the twelve control groups, 144 in all. It is on the basis of the consistency of the results of this large number of comparisons that the general trend concerning the susceptibility of different degrees of intelligence to retroaction assumes considerable reliability.

In some of the conditions of this part of the procedure the numbers appear small, particularly in certain of the dull groups. This is not because of insufficient subjects for the investigation as a whole, but because so many interpolation times were used that the cases were distributed over six different conditions and among four age-groups. In spite of the small sampling at the extreme age levels, however, the results are consistently in the same direction.

Scores on the initial test tended to vary with level of brightness. At both recalls, with all age levels, and in every period of interpolation the absolute losses sustained by the experimental groups were considerably greater than those of the control groups of the same degree of intelligence. Though there were appreciable differences in the percentages of retention of the bright, average, and dull, these were found chiefly in the experimental groups, the control groups differing but slightly in the percentage of words recalled. The differences in memory, corresponding to level of brightness, of the control groups were in the amounts learned rather than in the proportion of words recalled. In the experimental groups, with few exceptions, the percentage of retention of the average groups was greater than that of the dull; and, in every case, the retention of the bright surpassed that of both the average and the dull, in most instances by an appreciable amount.

To provide a single measure of retroaction for the bright, average, and dull levels, the percentages of inhibition for the four age levels were averaged. Table IV gives these averages at the seventeen-minute recall.

TABLE IV

AVERAGE PERCENTAGES OF RETROACTIVE INHIBITION FOR ALL AGES SEVENTEEN-MINUTE RECALL

| | Bright IQ Above 105 | Average 95–105 | Dull Below 95 |
|---|---|---|---|
| 4-Minute Interpolation | 26.9 | 34.8 | 39.2 |
| 6-Minute Interpolation | 29.3 | 38.0 | 44.6 |
| 8-Minute Interpolation | 34.0 | 40.0 | 42.5 |
| 10-Minute Interpolation | 33.8 | 45.7 | 53.5 |
| 12-Minute Interpolation | 34.1 | 46.7 | 49.3 |
| 14-Minute Interpolation | 32.3 | 53.4 | 57.5 |
| Mean | 31.7 | 43.1 | 47.8 |

These data indicate clearly that, at the seventeen minute recall, the amount of interference is consistently greater for the average than for the bright, and greater for the dull than for the average. In each of the three ability groups the percentage of retroaction increases as the length of time given to interpolated study is extended. This tendency is especially evident among the average, there being no exception at this level. In spite of some irregularities, the amount of interference evidenced by the dull increases considerably as the interpolation periods increase in duration. Among the bright there is an increment in the percentage of inhibition as

interpolated study increases from four to six and from six to eight minutes, but thereafter the figures remain approximately constant. In addition to the consistent tendency for retroactive inhibition to vary with brightness, it is interesting to note that, at this recall, extended interpolated study does not produce the additional interference among the bright that it does among the average and the dull.

In the twenty-four hour recall practically the same varying detrimental effects prevail.

TABLE V

AVERAGE PERCENTAGES OF RETROACTIVE INHIBITION FOR ALL AGES TWENTY-FOUR-HOUR RECALL

| | Bright IQ Above 105 | Average 95–105 | Dull Below 95 |
|---|---|---|---|
| 4-Minute Interpolation | 30.7 | 38.8 | 44.8 |
| 6-Minute Interpolation | 33.9 | 40.4 | 45.4 |
| 8-Minute Interpolation | 34.8 | 41.9 | 43.1 |
| 10-Minute Interpolation | 38.6 | 52.3 | 57.7 |
| 12-Minute Interpolation | 32.6 | 51.5 | 54.4 |
| 14-Minute Interpolation | 28.0 | 51.0 | 60.7 |
| Mean | 33.1 | 46.0 | 51.0 |

At this interval, the average amounts of inhibition show that extended interpolation up to ten minutes increases the amount of retroactive inhibition in the bright groups. Extending interpolated study beyond ten minutes, however, appears to decrease the retroactive effect at this level. In delayed recall retroaction also increases through ten minutes of interpolation with those of average brightness, though thereafter susceptibility to the disintegrating influence of intervening activity appears to remain almost constant. With the dull, however, allowing for some inconsistencies, the interference increases consistently from one interpolation to the next through fourteen minutes, the longest period of interpolated study employed in this study.

The general trend found by averaging the amounts of retroaction is also indicated in the percentages of the detrimental influence found in the individual age-groups. Though at each age level there are exceptions to this general tendency, in most instances they are statistically insignificant. Table VI gives the data for the seventeen-minute recall.

These data show that, at the eight and nine year level, in five of the six conditions the bright groups are affected less by the interference effect of the interpolated material than the average; and that, in every instance,

TABLE VI

PERCENTAGES OF RETROACTIVE INHIBITION OF BRIGHT, AVERAGE, AND DULL GROUPS SEVENTEEN-MINUTE INTERVAL

| C. A. | 8 and 9 Years | | | 10 and 11 Years | | | 12 and 13 Years | | | 14 and 15 Years | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bright | Average | Dull | Bright | Average | Dull | Bright | Average | Dull | Bright | Average | Dull |
| 4-Min. Int. | 25.9 | 32.6 | 37.5 | 28.8 | 40.0 | 41.7 | 28.9 | 37.7 | 42.2 | 24.0 | 29.0 | 35.5 |
| No. | 39 | 20 | 10 | 38 | 31 | 22 | 36 | 30 | 22 | 53 | 45 | 16 |
| 6-Min. Int. | 33.3 | 37.5 | 43.1 | 35.3 | 52.3 | 57.6 | 27.4 | 35.6 | 41.0 | 21.3 | 26.7 | 36.6 |
| No. | 26 | 26 | 10 | 22 | 39 | 36 | 44 | 37 | 32 | 40 | 35 | 21 |
| 8-Min. Int. | 35.5 | 39.5 | 47.5 | 36.8 | 47.4 | 45.2 | 36.1 | 39.8 | 40.5 | 27.7 | 33.3 | 36.7 |
| No. | 39 | 22 | 5 | 40 | 40 | 20 | 42 | 40 | 29 | 66 | 45 | 24 |
| 10-Min. Int. | 36.3 | 41.5 | 64.5 | 31.5 | 46.9 | 63.2 | 36.8 | 49.4 | 46.2 | 30.4 | 45.0 | 40.0 |
| No. | 49 | 32 | 7 | 36 | 58 | 27 | 66 | 49 | 22 | 57 | 40 | 15 |
| 12-Min. Int. | 39.6 | 38.6 | 56.4 | 37.9 | 58.3 | 57.1 | 38.1 | 47.7 | 42.7 | 20.8 | 42.2 | 41.1 |
| No. | 53 | 22 | 7 | 40 | 54 | 18 | 50 | 47 | 18 | 35 | 40 | 24 |
| 14-Min. Int. | 36.7 | 53.7 | 52.0 | 41.1 | 61.4 | 67.0 | 24.7 | 49.7 | 65.9 | 26.5 | 48.9 | 45.0 |
| No. | 65 | 29 | 11 | 59 | 63 | 32 | 58 | 60 | 16 | 34 | 44 | 11 |
| Mean | 34.6 | 40.6 | 50.2 | 35.2 | 51.1 | 55.3 | 32.0 | 43.3 | 46.4 | 25.1 | 37.5 | 39.2 |
| Control No. | 64 | 45 | 16 | 72 | 68 | 45 | 96 | 81 | 45 | 87 | 43 | 23 |

the dull suffer from the disintegrating influence of interpolation more than either the bright or the average. At the ten and eleven year level the same tendency appears, and, with two statistically insignificant exceptions, retroaction increases consistently with decreasing ability. The consistency of this tendency indicates the reliability of the conclusion that, at these age levels, there is an inverse relationship between degree of brightness and the percentage of retroactive inhibition that is produced by a similar serial list upon an antecedently learned list.

At the twelve and thirteen year level the difference between the average and the dull groups at the eight-minute interpolation is insignificant, while at the ten and twelve-minute interpolations it is actually in favor of the dull groups. Since the phenomenon seems to vary in its influence with individuals, the effect is evidenced with only certain of the subjects and with these in varying degrees. It is possible, therefore, that individual differences in susceptibility to the disturbing influence of intervening conditions have produced the fluctuations which appear to center at these ages. At the fourteen and fifteen year level, in the ten and fourteen-minute interpolations, the degree of retroaction is less with the dull than with the average groups, and at the twelve-minute interpolation the difference is insignificant. Here, also, the exceptions may be attributed to individual differences in susceptibility to the disintegrating effect of interpolated activity. Since the numbers in the groups in which the dull appear to surpass the average are smaller than in most of the conditions, the inconsistencies may be due, in part at least, to inadequate sampling.

It is particularly noteworthy that there is only one instance (the twelve-minute interpolation of the eight and nine year level) in which the retroactive influence in the bright groups exceeds that produced in the average and the dull. Equally as significant is the fact that at the seventeen-minute recall, though there are twenty-four conditions for comparison, there is only one case in which the percentage of inhibition of the bright groups (the fourteen-minute interpolation of the ten and eleven year age-level) reaches or exceeds the 40 per cent point, while, except in the shorter interpolation periods, the percentages of inhibition in the average groups are ordinarily much above 40 per cent. With the dull, except in the fourteen and fifteen year groups where there is a decided drop in the retroactive effect at all levels of ability, and in the four-minute interpolation of the

TABLE VII

PERCENTAGES OF RETROACTIVE INHIBITION OF BRIGHT, AVERAGE, AND DULL GROUPS
TWENTY-FOUR HOUR RECALL

| C. A. | 8 and 9 Years | | | 10 and 11 Years | | | 12 and 13 Years | | | 14 and 15 Years | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bright | Average | Dull | Bright | Average | Dull | Bright | Average | Dull | Bright | Average | Dull |
| 4-Min. Int. | 36.0 | 38.7 | 46.1 | 33.6 | 44.6 | 50.3 | 27.1 | 39.9 | 45.6 | 26.2 | 32.1 | 37.0 |
| Average I. Q. | 109 | 101 | 90 | 111 | 100 | 89 | 111 | 101 | 88 | 113 | 100 | 81 |
| 6-Min. Int. | 36.7 | 37.7 | 43.7 | 37.9 | 48.0 | 55.0 | 30.1 | 37.6 | 43.1 | 30.9 | 38.4 | 39.8 |
| Average I. Q. | 112 | 101 | 90 | 111 | 100 | 86 | 112 | 100 | 88 | 114 | 101 | 87 |
| 8-Min. Int. | 44.9 | 46.5 | 51.3 | 30.1 | 42.1 | 35.8 | 33.8 | 41.9 | 44.3 | 30.2 | 36.9 | 41.1 |
| Average I. Q. | 114 | 100 | 87 | 113 | 101 | 85 | 112 | 101 | 85 | 113 | 100 | 89 |
| 10-Min. Int. | 48.5 | 56.0 | 72.7 | 38.4 | 53.1 | 65.1 | 32.5 | 51.3 | 47.7 | 34.8 | 48.9 | 45.1 |
| Average I. Q. | 113 | 101 | 89 | 111 | 101 | 87 | 113 | 100 | 91 | 113 | 101 | 87 |
| 12-Min. Int. | 46.5 | 64.2 | 73.1 | 32.0 | 55.5 | 62.7 | 34.0 | 45.6 | 37.0 | 17.8 | 40.7 | 44.8 |
| Average I. Q. | 115 | 102 | 92 | 113 | 101 | 90 | 113 | 101 | 89 | 116 | 101 | 86 |
| 14-Min. Int. | 39.7 | 53.1 | 65.0 | 32.0 | 57.4 | 63.6 | 17.0 | 49.0 | 62.8 | 23.2 | 44.3 | 51.2 |
| Average I. Q. | 115 | 103 | 91 | 114 | 101 | 88 | 114 | 101 | 90 | 115 | 102 | 87 |
| Mean | 42.1 | 49.4 | 58.7 | 34.0 | 50.1 | 55.4 | 29.1 | 44.2 | 46.8 | 27.2 | 40.2 | 43.2 |
| Control Average I. Q. | 111 | 101 | 91 | 112 | 100 | 87 | 112 | 101 | 86 | 114 | 101 | 88 |

eight and nine year level, the amount of inhibition is considerably greater, varying from 40.5 per cent to 67.0 per cent.

When the individual conditions of the twenty-four hour recall are considered (Table VII), in nearly every comparison, though in some cases the differences are small, there appears a tendency for the percentage of inhibition to decrease in the order of dull, average, bright.

With these data as a basis for judging, it is evident that there can be but one conclusion, viz., that, under the conditions of this experiment, susceptibility to retroactive inhibition decreases with increasing degrees of brightness. When it is considered that one hundred forty-four comparisons were made possible by the conditions of the experiment, the consistency of the results argues for a high degree of reliability in the data. It is significant that most of the deviations from the general tendency occur in the two higher age levels where inconsistencies are also noted in the influence of the duration of the interpolated activity upon the degree of inhibition. With one exception, all deviations occur in the direction of the average and the dull groups. Only in the case of the twelve-minute interpolation of the eight and nine year level is there an exception among the bright groups. In this one case the amount of the difference in favor of the average group is so small as to be insignificant. In every other comparison the amount of retroaction is less among the bright than among the average. Allowing for some inconsistencies, there is also less inhibition among the average than among the dull. The difference in the inhibitory effect between the bright and the dull groups is decidedly significant in every condition. It is evident, therefore, that there is a definite inverse relationship between susceptibility to retroactive inhibition and degree of brightness.

From the consistency of the results obtained under the conditions of this experiment, the following conclusions appear justified.

1. Retroactive inhibition in children is a function of degree of intelligence and of the length of time devoted to interpolated study.

2. Chronological age being constant, susceptibility to the retroactive influence varies with level of brightness. The percentage of inhibition decreases as degree of brightness increases.

3. As indicated by the performance of the groups that had only four minutes of interpolation, the retroactive effect is at its maximum rate per minute during a comparatively short interval immediately following the original learning. The detrimental influence, however, increases slowly but regularly as the duration of the interpolated activity is augmented. This tendency is evident chiefly in the degree of retroaction of the average and dull groups.

4. Among the bright there is a tendency for the average percentage of retroactive inhibition to increase as the duration of the interpolated study is extended from four to eight or ten minutes; but thereafter the inhibitory effect remains practically constant.

5. In delayed recall, with average and bright groups, there appears to be an optimal interval beyond which increase in the duration of the period of interpolation produces no further increment in the retroactive influence. The same trend is evident, to a lesser degree, in the dull groups. With the bright there is a marked tendency for the degree of inhibition to decrease as the period of interpolation extends beyond ten minutes.

6. With brightness held constant, susceptibility to retroaction decreases with increasing chronological and/or mental age, the results at the eight and nine year level constituting an exception that may be attributed to small sampling and to the low scores attained on the initial test.

7. Since the bright, average, and dull control groups differ in original learning scores, but not appreciably in comparative retention, variability in the retroactive effect may be traced almost entirely to the influence of the interpolated activity.

8. The degree of retroactive inhibition varies considerably with individuals of all age levels and of all degrees of brightness, though not sufficiently to conceal strong group tendencies.

# THE CONDITIONED REFLEX TECHNIQUE APPLIED TO A LESS SPECIALIZED TYPE OF LEARNING

MERRILL T. EATON
*Indiana University*

Much has been written concerning the conditioned reflex, and numerous extensive experimental investigations have been conducted in order to determine the methods and techniques by which the reflex may most efficiently be conditioned. As a result of these studies, the conditioned reflex technique has been established as an important method of modifying certain types of behavior.

Although many aspects of the conditioned reflex technique were considered before Pavlov's work, the theory of conditioned learning was developed largely as the result of his studies. This theory (14) attempts to explain all modifications of behavior on the basis of the conditioned reflex. Allport's (1) theory of the development of social behavior, Watson's (20) behaviorism, and Smith and Guthrie's (16) work are typical of this concept.

If the conditioned reflex theory of learning is applicable to all kinds of everyday learning, the conditioned reflex technique applied to a less specialized type of learning should yield valuable data concerning the learning process.

A survey of the literature concerning the applicability of the conditioned reflex theory to all modifications of behavior indicates a scarcity of experimental studies which have definitely applied the conditioned reflex technique to the less specialized types of learning. Studies such as those of Jones (11) and of Watson and Rayner (21) do, however, in a general way demonstrate the application of the conditioned reflex technique in modifying emotional behavior. Thorndike (18), in his investigation of what he calls "the principle of belongingness," seems very definitely to have used a phase of the conditioned reflex technique.

There also seems to be some uncertainty among students of the conditioned reflex phenomena as to whether all of the so-called conditioned reflexes are really true reflexes. Tolman (19) limits conditioned reflex learning to the conditioning of the true reflex; i.e., attaching of the original unconditioned response to the conditioned stimulus. Other aspects of the conditioning process, which are less specialized than the true conditioned reflex, have been called "sign learning" by Tolman. "Sign learning," according to Tolman, consists of the everyday type of learning in which one learns to respond to a stimulus in a different way from that in which he first responded to it. This interpretation seems to be supported by Cason (5), who says there is considerable doubt whether all the conditioned reactions which have been secured were true conditioned reflexes.

### *The Conditioned Reflex Technique*

In order to establish a conditioned response there must be a stimulus capable of eliciting a response (the unconditioned stimulus) and another stimulus to which the response is not attached (the conditioned stimulus), and these two stimuli must be presented in rather close temporal relationship. The conditioned response is established when the inappropriate stimulus is substituted for the unconditioned stimulus. It is generally considered that in the everyday type of learning either side (the S or the R) of the response mechanism may be conditioned.

An important aspect of the conditioned reflex technique is the method of presenting the stimulus relative to time relationship. The conditioned stimulus may be simultaneous with, may precede, or may follow the unconditioned stimulus. In this investigation the following names are used to distinguish the three time relationships:

1. Simultaneous. The conditioned stimulus and the unconditioned stimulus are presented simultaneously.
2. Delayed. The conditioned stimulus precedes the unconditioned. The two do not overlap in time.
3. Backward. The conditioned stimulus follows the unconditioned. The two do not overlap in time.

The conditioned reflex theory assumes that all learning reduces to conditioned reflexes. If the conditioned reflex technique, as described, is applied to a simple, unspecialized type of learning comparable with everyday learning situations, as contrasted to the highly specialized reflex, a number of questions pertinent to learning efficiency are raised:

1. Does the modification of less specialized behavior by the conditioned reflex technique follow the same laws as does the conditioning of true reflexes?
2. Which of the three methods of presenting the terms (simultaneous, delayed, or backward) is most efficient for learning?
3. What effect does the length of the time interval between the presentation of terms have upon the efficiency of learning?
4. Will learning efficiency be affected if the time intervals are filled with interpolated activities?
5. Will learning be facilitated by certain kinds of interpolated activities and be retarded by others?
6. By which method of presenting terms (simultaneous, delayed, or backward) will learning be most stable (remembered best)?
7. Are modifications of less specialized types of behavior by the conditioned reflex technique unstable, and are these modifications soon lost?
8. Will retention be affected by interpolated activities?

The purpose of this investigation is to attempt to answer the above questions by applying the conditioned reflex technique to a simple type of learning.

The operation of the conditioned reflex technique in modifying the true reflexes seems to be relatively well established. Pavlov (14) found in working with the salivary reflex that the simultaneous method of presenting stimuli was most effective and that the length of the time interval between the unconditioned and the conditioned stimulus was a determining factor in the ease of conditioning. He also discovered that if the conditioned stimulus preceded the unconditioned, conditioning was extremely difficult, and that modifications of reflex behavior by conditioning were unstable. The results of more recent investigation of the conditioned reflex (10) indicate that the results of conditioning may be relatively stable. Wolfle (22), investigating the reflex jerk of the finger, found backward conditioning difficult. She also found that the length of the time interval between terms was important in ease of conditioning.

Froeberg (7), working with nonsense syllable pairs, found that the longer time interval and interpolated activities between terms reduced learning. Bergstrom's (3) results in a somewhat similar study would indicate that the optimal interval between terms depends at least in part upon the type of material to be learned. Perhaps the most important of all studies dealing with the effect of time in learning is that of Snoddy (17). In his very elaborate study, Snoddy showed that time in relation to maturation (experience) must be considered as an important factor.

The theory has been advanced that association never takes place between terms unless they are presented simultaneously. The proponents of this theory insist that where associations do take place between terms separated by a time interval, it is because of the fact that, when the second term is presented, the verbal, mental, motor, or kinaesthetic image of the first is reproduced; hence it reduces to simultaneity of presentation. The soundness of this theory seems not to have been definitely demonstrated by experimental investigations.

One can find much relevant material concerning the problem of interpolated activities between learning and recall in studies dealing with the phenomena of retention. Studies which involved interpolated activities which affected learning and retention, "retroactive inhibition," and "perseveration" are rather numerous. One of the earlier studies which was definitely concerned with the problem of retroactive inhibition and perseveration was that of Müller and Pilzecker (13). Their investigations, using nonsense syllables as the material to be learned and retained, with the interpolated activities between learning and recall being rest, memorizing a second list of syllables, or memorizing pictures, found that the perseverative process was most influential immediately after learning and then gradually lost its effect. Heine (9) and De Camp (6), investigating the same problems, secured conflicting results. Heine's conclusions were in

general agreement with those of Müller and Pilzecker; but De Camp found little difference in results when an interpolated activity of multiplication was compared with a rest period.

The most extensive study of the effect of the similarity of interpolated activities on the retention of the original materials is that of Robinson (13). A number of interpolated activities were inserted between learning and recall, varying from very similar tasks to those which were quite dissimilar. Robinson's results showed that the highly similar interpolated tasks affected learning efficiency most and those which were least similar affected learning least.

*Learning Materials*

After considerable experimentation with different types of materials, 30 pictures of objects and people, approximately 7″ x 10″ in size, were selected as being the most satisfactory type of learning materials. The pictures were mounted on 9″ x 12″ drawing paper and each picture was given a two-word name. There was no connection, either in current use, or of the elements in the picture, between the name and the picture. By random selection the 30 pictures were divided into three groups of 10 each. The three groups of pictures were labeled respectively simultaneous, delayed, and backward. Each picture was rotated through the three methods of presentation so that any differences in ease of learning would be equally distributed. These pictures were named the major pictures in order to distinguish them from the interpolated minor pictures.

Five different time intervals (20, 30, 40, 50, and 60 seconds) between the presentation of terms were studied; and five different activities (rest, arithmetic, pied type, names, and pictures) were interpolated during the time intervals. The different learning situations are shown in Table I. The effect of interpolated activities on learning efficiency was studied for all the time intervals except interpolated pictures during the 20-second interval. The numbers in Table I refer to the number of subjects for each time interval in each learning situation.

The interpolated activities varied from highly dissimilar tasks to tasks which were quite similar to the major learning. The arithmetic problems included addition, subtraction, multiplication, and division; the pied type was single-spaced; each name consisted of two words and was quite similar to, although not the same as, the names of the major pictures; the minor pictures were approximately one-half the sixe of the major pictures and were mounted on drawing paper. The minor pictures were given names consisting of two words and the name was descriptive of the picture.

*Experimental Procedure*

The conditioned reflex technique was thoroughly explained to the subject before the learning materials were presented. The subjects were definitely informed of just what they were expected to learn and how the efficiency of their learning would be tested. In order to be sure that the explanations were fully understood, materials similar to the learning materials were used to demonstrate each phase of the learning situation. The laboratory in which the experiment was conducted consisted of a normal, natural learning environment, and not more than seven subjects were involved at one time.

At the beginning of each learning situation the following additional information and directions were given the subjects: "You will be shown 30 pictures. Each picture will be presented for a period of 10 seconds and its

TABLE I

LEARNING SITUATIONS—IMMEDIATE TEST GROUP

| Length of interval in seconds | (Rest) Unfilled | Filled Arithmetic | Filled Pied Type | Filled Names | Filled Pictures |
|---|---|---|---|---|---|
| 20 | 23 | 14 | 21 | 20 | |
| 30 | 27 | 21 | 14 | 20 | 14 |
| 40 | 19 | 26 | 37 | 16 | 12 |
| 50 | 24 | 21 | 20 | 15 | 30 |
| 60 | 40 | 24 | 34 | 14 | 14 |
| N | 133 | 106 | 126 | 85 | 70 |

name will be distinctly spoken twice. There will be a 10-second interval between each completed unit (picture-name) of the picture series and the next unit. For 10 of these pictures, the picture and the name will be given at the same time; for 10 the name will be given a unit of time preceding the picture; and for 10 the name will be given a unit of time after the picture has been shown. The order in which the names and pictures are presented, however, will be an entirely chance order. Your task is to learn and to remember the correct name for each picture so that when the pictures are again presented you can give their correct names."

The method of testing learning efficiency was the same for all time intervals in all learning situations. As soon as the last of the learning terms had been presented sheets were handed to the subjects with the numbers from 1 to 30. Exactly three minutes after the last learning term had been presented the pictures were shuffled and each picture was presented for 10 seconds; but this time only the number of the picture was given, and the subjects were asked to write the name of the picture following the correct number on their sheets. Immediately following this test, if the interpolated activities were minor pictures or names, the subjects were tested on these activities.[1] Learning efficiency was measured on the basis of the number of correct responses. If both words in the name of the major picture were given correctly, it was scored two points; if one word of the name was given correctly, it was scored one point.

One group of subjects, immediate test group, ($G_1$), was given only the one test, immediately following the learning period. The other group, re-test group, ($G_2$), was given in addition to the immediate test, four memory tests at intervals of one week, one week, one week, and two weeks. At no time during this five-week period did the subjects have any opportunity to see the major pictures or to review the learning materials in any way. Neither were the subjects of $G_2$ informed that they were to have any of the memory tests. The memory tests were conducted in the same manner as the immediate test except that for each re-test the numbers of the pictures were changed.

[1] The performance of the subjects on the interpolated activities was analyzed but the data have little bearing on the major problems of this study and are too voluminous to be included.

The experimental procedure for presenting the terms is illustrated in Figure 1. The diagram shows the plan of presentation of the major pictures and the interpolated arithmetic problems. The plan shown is for the 20-second interval, but is representative of the other time intervals in the series, except that the interval between the words "work" and "stop" would be 30, 40, 50, or 60 seconds. If the term "rest", "pied type", "names", or "pictures" were substituted for the word "problems", this diagram would be representative of the other four learning situations.

During the interval between the commands "work" and "stop", the subjects were instructed as follows: (1) unfilled—relax as completely as possible; (2) arithmetic—work continuously on the problems, both speed and accuracy are important; (3) pied type—work steadily at checking the designed letter, both speed and accuracy are important; (4) names—give close attention to the names read to you; you will be tested on the number you remember; for the 20-second interval there will be one name, two for the 30, three for the 40, four for the 50, and five names for the 60-second interval; (5) pictures—give close attention to the minor pictures, since they will be presented again and you will be expected to remember their names; for the 30-second interval there will be one minor picture, two for the 40, three for the 50, and four during the 60-second interval. At the command "stop", give your undivided attention to the experimenter.

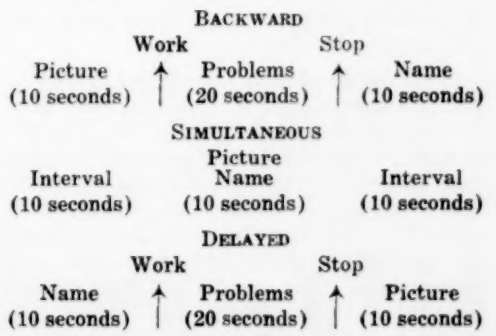



**Figure 1.—Plan of presentation of major pictures and interpolated activities.**

*Subjects*

The subjects were 608 college students. There were 520 subjects in the immediate test group, and 88 subjects in the re-test

TABLE II

AGE, SEX, AND CLASS DISTRIBUTION OF SUBJECTS—G1

| Age | | | | | | | | Sex | | Class | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23–28 | M | F | F | S | Jr | Sr | N |
| 6 | 36 | 120 | 157 | 118 | 47 | 16 | 20 | 286 | 234 | 265 | 203 | 35 | 17 | 520 |

TABLE III

AGE, SEX, AND CLASS DISTRIBUTION OF SUBJECTS—G2

| Age | | | | | | | Sex | | Class | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 18 | 19 | 20 | 21 | 22 | 23–40 | M | F | F | S | Jr | Sr | N |
| 10 | 17 | 25 | 16 | 7 | 3 | 10 | 50 | 38 | 67 | 11 | 7 | 3 | 88 |

group. The age, sex, and class standing of the two groups are given in Tables II and III. The subjects participating in each time interval in each learning situation were unselected. A survey of their percentile ratings,[2] which was the only available measure of their ability (intelligence), indicated that the average ability of the different groups of subjects was approximately equal. The subjects in the immediate test groups, however, were on the average more mature both in age and class standing than were the subjects of the re-test groups.

## RESULTS

Immediate Test Group.—Modifications in learning efficiency resulting from the following three factors were studied in this phase of the investigation: (1) the method of presenting the terms, simultaneous, delayed or backward; (2) the length of the time interval between terms; and (3) the nature of the interpolated activities.

(1) *The Method of Presenting Terms.*—The average[3] scores for each of the three methods of presenting terms for the 520 subjects of the immediate test group for the five time intervals in the five learning situations were computed. These data are given in Table IV. The score made by the simultaneous method of presenting terms was 2.22 higher than that made by the delayed method; the delayed was 1.90 higher than the backward; and the simultaneous was 4.12 higher than the backward. It will be noted that these differences between the average scores are considerably higher than necessary for complete statistical reliability.[4]

[2] Psychological Examination for High School Graduates and College Freshmen, prepared by L. L. and T. G. Thurstone, published by the American Council on Education.

[3] Due to the voluminous character of the individual data only the average scores for each time interval in each learning situation are presented.

TABLE IV

THE AVERAGE SCORES FOR EACH OF THE THREE TIME RELATIONSHIPS, S, D, AND B MADE BY THE 520 SUBJECTS—G1

| | S | D | B | S |
|---|---|---|---|---|
| Ave. ------ | 13.10 | 10.88 | 8.98 | 13.10 |
| Diff. ------ | 2.22 | 1.90 | 4.12 | |
| Rel. ------ | 6.94 | 5.43 | 12.87 | |

A further comparison of the extent of learning for each of the three methods is presented in Figure 2. The solid bar represents the average score made by the simul-

[4] The length of tables has been reduced by omission of all intermediate statistical steps. Reliability of the differences between averages is in terms of the sigma (diff.) (8). A $\frac{D}{\text{sigma diff.}}$ of 3 indicates practically complete reliability.

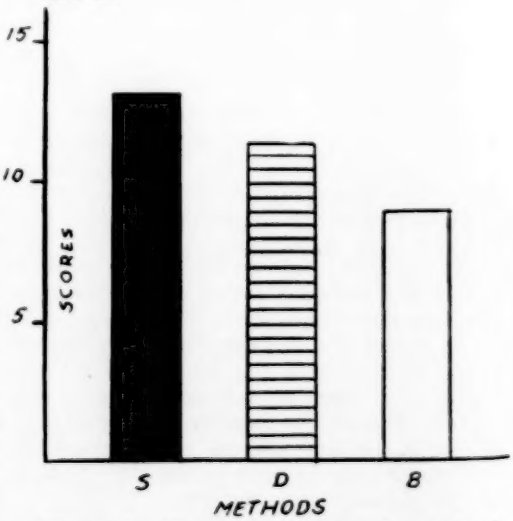



Figure 2.—The average scores for each of the three time relationships, S, D and B, made by the 520 subjects—G1.

taneous method of presenting terms; the lined bar the delayed; and the open bar the backward. This graph shows plainly that, when the composite scores for the three methods of presenting terms are considered, the simultaneous method is far more efficient than the delayed or backward, and that the delayed method is superior to the backward.

(2) *The length of the time interval between the presentation of terms.*—Variations in learning by the S, D, and B method of presenting terms resulting from the length of the time interval may be studied by a comparison of the S, D, and B scores for the different time intervals in the unfilled learning situation. These scores are given under the caption "unfilled" in Table V. There is little difference between the average S, D, and B scores in the 20-second interval. The differences that do exist possess such low reliability that they may be ignored. The average S and D scores for the 30-second interval are approximately equal but the B score is considerably lower. The differences between the average scores of D and B (3.37) and between B and S (3.52) are reliable differences. The S and D scores are higher for the 30-second interval than they were for the 20-second, but the B score is lower.

The average S, D, and B scores are higher for the 40-second interval than they were for either the 30 or 20-second interval. There is little difference in the S and D scores, but the B score is 2.11 lower than D and 2.69 lower than S. The S, D, and B scores are widely separated for the 50-second interval and lower than they were for the 40-second interval. For the 60-second interval the S and D scores are approximately equal, but the B score is 2.55 lower. The S score is lower for the 60 than it was for the 50-second interval but the D and B scores are higher. The effect of the length of the time interval upon learning efficiency by the S, D, and B method of presenting terms may be readily seen if the curves (A unfilled) in Figure 3 are examined.

The total average scores (the sum of the S, D, and B scores) for each time interval are given under the caption "unfilled" in Table VI. When these total average scores for each time interval in the unfilled learning situation are examined it is immediately noted that *the shortest time interval between terms is not conducive to the most efficient learning*. The total average score for the 20-second interval is 40.17, while for the 30-second interval it has increased to 45.56.

TABLE V

THE AVERAGE SCORES MADE BY THE SIMULTANEOUS, DELAYED, AND BACKWARD METHODS OF PRESENTING TERMS FOR EACH TIME INTERVAL IN EACH LEARNING SITUATION—G1

| | 20 Seconds | | | 30 Seconds | | | 40 Seconds | | | 50 Seconds | | | 60 Seconds | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ave. | Diff. | Rel. | Ave. | Diff. | Rel. | Ave. | Diff. | Rel. | Ave. | Diff. | Rel. | Ave. | Diff. | Rel. |
| UNFILLED | | | | | | | | | | | | | | | |
| S | 13.17 | | | 16.41 | | | 16.95 | | | 15.63 | | | 13.43 | | |
| D | 13.43 | .26 | .18 | 16.26 | .15 | .19 | 16.37 | .58 | .48 | 11.71 | 3.92 | 2.90 | 14.33 | .90 | .95 |
| B | 13.57 | .14 | .11 | 12.89 | 3.37 | 3.21 | 14.26 | 2.11 | 1.45 | 9.67 | 2.04 | 1.38 | 10.88 | 3.45 | 3.11 |
| S | 13.17 | .40 | .10 | 16.41 | 3.52 | 3.06 | 16.95 | 2.69 | 1.33 | 15.63 | 5.96 | 4.19 | 13.43 | 2.55 | 2.50 |
| NAMES | | | | | | | | | | | | | | | |
| S | 11.85 | | | 13.70 | | | 15.06 | | | 16.40 | | | 16.43 | | |
| D | 9.60 | 2.25 | 1.36 | 11.95 | 1.75 | .88 | 12.06 | 3.00 | 2.31 | 15.47 | .93 | .65 | 11.86 | 4.57 | 3.01 |
| B | 8.20 | 1.40 | .82 | 11.85 | .10 | .05 | 10.63 | 1.43 | .99 | 13.33 | 2.14 | 1.30 | 12.14 | .28 | .17 |
| S | 11.85 | 3.65 | 2.05 | 13.70 | 1.85 | 1.15 | 15.06 | 4.43 | 3.10 | 16.40 | 3.07 | 2.26 | 16.43 | 4.29 | 2.80 |
| PIED TYPE | | | | | | | | | | | | | | | |
| S | 11.00 | | | 13.86 | | | 12.11 | | | 15.75 | | | 12.76 | | |
| D | 10.48 | .52 | .39 | 13.79 | .07 | .05 | 11.68 | .43 | .43 | 8.85 | 6.90 | 5.89 | 12.38 | .38 | .38 |
| B | 7.71 | 2.77 | 2.18 | 11.07 | 2.72 | 1.94 | 9.14 | 2.54 | 2.39 | 5.65 | 3.20 | 2.32 | 7.79 | 4.59 | 4.73 |
| S | 11.00 | 3.29 | 2.91 | 13.86 | 2.79 | 1.99 | 12.11 | 2.97 | 2.91 | 15.75 | 10.10 | 8.94 | 12.76 | 4.97 | 4.89 |
| PICTURES | | | | | | | | | | | | | | | |
| S | | | | 13.21 | | | 13.25 | | | 13.93 | | | 10.79 | | |
| D | | | | 10.71 | 2.50 | 1.77 | 8.17 | 5.08 | 2.54 | 8.83 | 5.10 | 4.05 | 4.79 | 6.00 | 3.47 |
| B | | | | 9.71 | 1.00 | .76 | 8.92 | .75 | .35 | 7.73 | 1.10 | .79 | 5.93 | 1.14 | .69 |
| S | | | | 13.21 | 3.50 | 2.40 | 13.25 | 4.33 | 2.38 | 13.93 | 6.20 | 5.49 | 10.79 | 4.86 | 2.68 |
| ARITHMETIC | | | | | | | | | | | | | | | |
| S | 10.57 | | | 9.19 | | | 10.08 | | | 8.43 | | | 10.38 | | |
| D | 11.64 | 1.07 | 1.02 | 7.95 | 1.24 | .88 | 5.73 | 4.35 | 3.45 | 5.14 | 3.29 | 2.26 | 7.96 | 2.42 | 1.65 |
| B | 7.29 | 4.35 | 3.30 | 4.67 | 3.28 | 2.65 | 3.08 | 2.65 | 2.26 | 3.38 | 1.76 | 1.43 | 6.08 | 1.88 | 1.48 |
| S | 10.57 | 3.28 | 2.73 | 9.19 | 4.52 | 3.93 | 10.08 | 7.00 | 6.60 | 8.43 | 5.05 | 3.80 | 10.38 | 4.30 | 3.23 |

There is a slight additional increase up to 47.58 for the 40-second interval. For the 50-second interval the score drops to approximately the same as for the 20-second interval and maintains the same relative position although the trend is upward for the 60-second interval. For the unfilled intervals, the 30- and 40-second intervals lead to the most efficient learning. A glance at the statistical reliability (Table VI) of the differences between the total average scores shows that they possess high reliability. The marked increase in learning efficiency with increase in the length of the time interval is shown by the curve (F-totals) marked "unfilled" in Figure 3.

Time intervals of different lengths between the presentation of terms operated in two distinct ways to produce variations in learning: (1) changes in the average scores made by the S, D, and B technique; and (2) increase in the length of the time interval through the 40-second interval increased learning.

(3) *Interpolated activities*—Variations in learning efficiency by the S, D, and B technique resulting from the interpolated activities may be studied by comparing the S, D, and B scores for each time interval in the filled learning situations with those of the unfilled.

In all but three instances, the 20 and 60-second intervals in the unfilled learning situa-

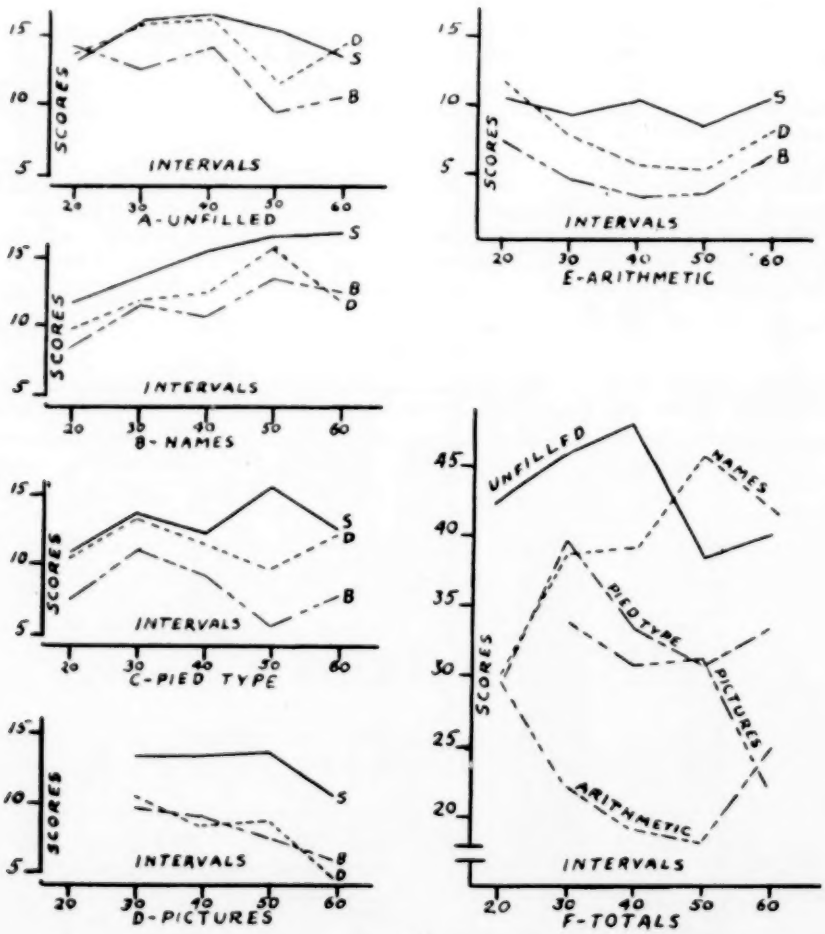



Figure 3.—The efficiency of learning by the S, D, and B methods of presenting terms.

TABLE VI

THE TOTAL AVERAGE SCORES FOR EACH TIME INTERVAL IN EACH LEARNING SITUATION—G1

| | 20 Sec. | | 30 Sec. | | 40 Sec. | | 50 Sec. | | 60 Sec. | | 20 Sec. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| UNFILLED | | | | | | | | | | | |
| Ave. | 40.17 | | 45.56 | | 47.58 | | 37.00 | | 38.63 | | 40.17 |
| Diff. | | 4.39 | | 2.02 | | 10.58 | | 1.63 | | 1.41 | |
| Rel. | | 1.38 | | .58 | | 3.07 | | .51 | | .43 | |
| NAMES | | | | | | | | | | | |
| Ave. | 29.65 | | 37.50 | | 37.75 | | 45.20 | | 40.43 | | 29.65 |
| Diff. | | 7.85 | | .25 | | 7.45 | | 4.77 | | 10.78 | |
| Rel. | | 1.58 | | .06 | | 2.09 | | 1.22 | | 2.45 | |
| PIED TYPE | | | | | | | | | | | |
| Ave. | 29.19 | | 38.71 | | 32.92 | | 30.25 | | 32.94 | | 29.19 |
| Diff. | | 9.52 | | 5.79 | | 2.67 | | 2.69 | | 3.75 | |
| Rel. | | 2.63 | | 1.70 | | .95 | | .95 | | 1.21 | |
| PICTURES | | | | | | | | | | | |
| Ave. | | | 33.64 | | 30.33 | | 30.50 | | 21.50 | | |
| Diff. | | | | 3.31 | | .17 | | 9.00 | | | |
| Rel. | | | | .80 | | .05 | | 2.16 | | | |
| ARITHMETIC | | | | | | | | | | | |
| Ave. | 29.50 | | 21.81 | | 18.88 | | 16.95 | | 24.42 | | 29.50 |
| Diff. | | 7.69 | | 2.93 | | 1.93 | | 7.42 | | 5.08 | |
| Rel. | | 2.66 | | 1.00 | | .57 | | 2.06 | | 1.61 | |

tion and the 20-second interval in the interpolated arithmetic problems, learning was more efficient by the simultaneous method than it was by the delayed or backward. In only four cases are the scores higher by the backward method than by the delayed: the 20-second unfilled, 60-second interpolated names, and the 40- and 60-second interpolated pictures. Although these average scores are slightly higher, the statistical reliability of the differences (Table V) is only slightly better than chance. There is one instance, a 20-second unfilled, where the average score was higher by the backward method than by the simultaneous. The difference (Table V) is hardly more than a chance difference. When the average scores made by the simultaneous method are compared with those made by the backward method, the higher average scores are registered by the simultaneous method, except in the case mentioned, and in general the differences between the average scores made by the simultaneous and backward methods possess consistently high statistical reliability.

A comparison of the S, D, and B curves for the filled learning situations with those of the unfilled (Figure 3) shows that, although the interpolated activities did produce variations in the S, D, and B scores, the variations in the four filled learning situations were neither uniform nor constant. The efficiency of the S, D, and B technique was markedly changed by some of the interpolated activities, while other activities produced relatively small changes.

Changes in learning efficiency during each time interval resulting from the interpolated activities may be more clearly observed by comparing the total average scores (Table VI) for each time interval in the filled learning situations with those in the unfilled. The general trend of the scores in the interpolated name learning situation is similar to that in the unfilled.

In interpolated names, however, learning was most efficient during the 50-second interval; and, even for the 60-second interval, the score was well above the 30- and 40-second interval performance and far above the 20-second interval. The differences between these averages possess high statistical reliability.

In the interpolated pied type learning situation the general trend of the data has changed. In this case the 30-second interval was conducive to the most efficient learning, the scores on the other four intervals being rather close together. More changes are noted in the trend of the scores in interpolated pictures. In this learning situation the shortest time interval, the 20-second, produced the most efficient learning. There was a slight drop in the score from the 30- to the

40-second; the 50-second remained practically the same; but a considerable drop in score took place from the 50- to 60-second interval. The reliability of these differences is low, except the difference between the scores for the 50- and 60-second interval.

Very marked changes are noted in the scores for interpolated arithmetic. Here the shortest, 20-second, and the longest, 60-second, intervals produced the highest scores. The total average scores for the intermediate time intervals, 30, 40, 50 seconds, are considerably lower than for the short and long intervals.

Comparing the scores for the different time intervals in each learning situation, it is clear that the interpolated activities have produced a complete reversal in the trend of the scores in relation to the length of the time interval. This condition may be seen by examining the curves (F-totals) in Figure 3. The plotted scores for the unfilled time intervals give an almost perfect bell-shaped curve, showing the highest scores for the middle time intervals and the lowest scores for the short and long intervals. The curves for interpolated names and pied type are still bell-shaped, but not as regular as for the unfilled scores. For interpolated pictures the curve has attained a new shape. The curve for interpolated arithmetic is again bell-shaped but *completely inverted. From the unfilled scores through the interpolated activities the curves change from an upright bell-shape to an inverted bell-shape.*

The data in Table VI have been reorganized in Table VII so that the total average scores made in each time interval in the filled learning situations may be compared directly with the scores in the unfilled. Learning was most efficient for the unfilled learning situation. When activities were interpolated, names had the least deteriorative effect, followed by an increasing reduction in total average scores through pied type, pictures, and arithmetic.

The total average score for the 20-second unfilled interval is considerably higher than that for names. There is little difference, however, in the scores for names, pied type, and arithmetic. The differences of 10.52 between the scores for unfilled and names, and of 10.67 between unfilled and arithmetic possess high reliability.

For the 30-second interval the difference between the total average scores for unfilled and names is 8.06, which is near complete statistical reliability. The total average scores for names and pied type is about the same, but for interpolated pictures there is a drop in score of 5.07 below the score for pied type. Interpolated arithmetic caused another drop of 11.83 below the score on pictures. The differences between the average scores

TABLE VII

THE RELIABILITY OF THE DIFFERENCES BETWEEN THE TOTAL AVERAGE SCORES FOR EACH TIME INTERVAL IN EACH LEARNING SITUATION

| | Unfilled | | Names | | Pied type | | Pictures | | Arithmetic | | Unfilled |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 SECONDS | | | | | | | | | | | |
| Ave. | 40.17 | | 29.65 | | 29.19 | | | | 29.50 | | 40.17 |
| Diff. | | 10.52 | | .46 | | | | | | 10.67 | |
| Rel. | | 2.04 | | .11 | | | | | | 2.84 | |
| 30 SECONDS | | | | | | | | | | | |
| Ave. | 45.56 | | 37.50 | | 38.71 | | 33.64 | | 21.81 | | 45.56 |
| Diff. | | 8.06 | | 1.21 | | 5.07 | | 11.83 | | 23.75 | |
| Rel. | | 2.03 | | .26 | | 1.28 | | 3.33 | | 7.71 | |
| 40 SECONDS | | | | | | | | | | | |
| Ave. | 47.58 | | 37.75 | | 32.92 | | 30.33 | | 18.88 | | 47.58 |
| Diff. | | 9.83 | | 4.83 | | 2.59 | | 11.45 | | 28.70 | |
| Rel. | | 2.78 | | 1.50 | | .58 | | 2.57 | | 8.92 | |
| 50 SECONDS | | | | | | | | | | | |
| Ave. | 37.00 | | 45.20 | | 30.25 | | 30.50 | | 16.95 | | 37.00 |
| Diff. | | 8.20 | | 14.95 | | .25 | | 13.55 | | 20.05 | |
| Rel. | | 2.36 | | 4.66 | | .08 | | 3.85 | | 5.63 | |
| 60 SECONDS | | | | | | | | | | | |
| Ave. | 38.63 | | 40.43 | | 32.94 | | 21.50 | | 24.42 | | 38.63 |
| Diff. | | 1.80 | | 7.49 | | 11.44 | | 2.92 | | 14.21 | |
| Rel. | | .49 | | 2.08 | | 2.34 | | .57 | | 4.37 | |

for pictures and arithmetic, and arithmetic and unfilled, are reliable statistical differences.

For the 40-second interval the score for the unfilled learning situation is 9.83 higher than the score for interpolated names. This difference has high statistical reliability. There is also a drop of 4.83 between the scores of names and pied type. There is only a slight drop in score from pied type to pictures, but there is a difference of 11.45 between the scores of pictures and arithmetic. Between the unfilled and the interpolated arithmetic scores there is a difference of 28.70. This difference has full statistical reliability.

The total average score for interpolated names in the 50-second interval is 8.20 higher than the total average score for the unfilled interval. It will be observed that the difference between the scores for unfilled and interpolated names has high statistical reliability. The total average score for interpolated pied type is 14.95 lower than for names. This difference is completely reliable. There is very little difference between the scores for pied type and pictures, but for interpolated arithmetic the score is 13.55 lower than the interpolated picture score. This difference is completely reliable, as is the difference of 20.05 between the scores on interpolated rest and arithmetic. The score for the 60-second interval in interpolated names is 1.80 better than the unfilled learning situation. This difference is only slightly better than chance. The difference of 7.49 between the scores on interpolated names and pied type possesses high statistical reliability. There is a difference of 11.44 between the scores on interpolated pied type and pictures, and this difference also possesses high statistical reliability. There is a difference of only 2.92 between the score for interpolated pictures and arithmetic, but there is a difference of 14.21 between the scores on interpolated arithmetic and unfilled which is a reliable statistical difference.

The reduction in learning efficiency resulting from the different interpolated activities can be seen in the downward slope of the curve in Figure 4 from unfilled through the other interpolated activities down to arithmetic. The curve is plotted from the average score made in each learning situation. No sudden or extreme drops appear in the curve, but rather a slow, gradual decline.

This is an interesting phase of the study; other investigations have indicated that with interpolated activities learning would be reduced in proportion to the similarity of the interpolated activities to the major learning materials. Apparently this is not true in this study, since names and minor pictures are both quite similar to the major pictures, and pied type and arithmetic are quite dissimilar, yet learning was reduced least by names and most by arithmetic. This would indicate that the amount of concentration (attention) demanded by the interpolated activities was the dominant factor, rather than the similarity or lack of similarity of the interpolated activities to the major learning materials.

The interpolation of activities between the presentation of terms affected learning efficiency in three major respects: (1) produced changes in the relationship of the S, D, and B scores; (2) caused variations in learning in the different time intervals; and (3) caused a reduction in learning.

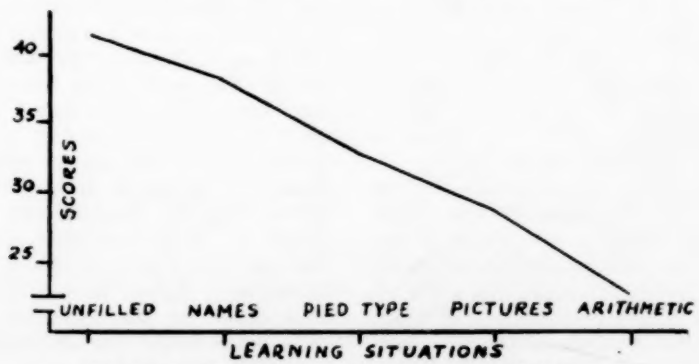



**Figure 4.—The effect of interpolated activities on learning efficiency.**

*Re-test group*

The 88 subjects in the re-test group were given, in addition to the test immediately following the presentation of the major pictures, four subsequent tests at intervals of one week between each of the first three tests and two weeks preceding the last test, thus distributing the tests over a period of five weeks. The experimental procedures for the re-test groups were identical with those of the immediate test groups, except that only one time interval, the 30-second, was studied.

The important problems involved in the study of the average scores made by the re-test group of subjects are: (1) By which method of presenting terms, S, D, or B, is learning retained best? (2) the variations in retention for the different tests; (3) variations in retention in relation to the interpolated activities; (4) a comparison of the efficiency of learning of the re-test group with that of the immediate test group.

(1) By which method of presenting terms, S, D, or B, is learning retained best? The average scores made by the S, D, and B methods for each learning situation in each test are given in Table VIII. Variations in retention by the S, D, and B method of presenting terms are somewhat more pronounced in interpolated arithmetic, pictures, and names than in the unfilled and pied type learning situations. The curves in Figure 5 show no marked variations in degree of retention by the three methods of presenting terms.

The curves in Figure 5 are plotted from the data given under averages (the averages S, D, and B scores for each test) at the bottom of Table VIII. These curves show that, although there are slight variations in retention from one test to another, retention in relation to the method of presenting terms is relatively constant. More is learned by the simultaneous method than by the delayed or backward, and more by the delayed than by the backward; but the degree of retention for each method is approximately the same.

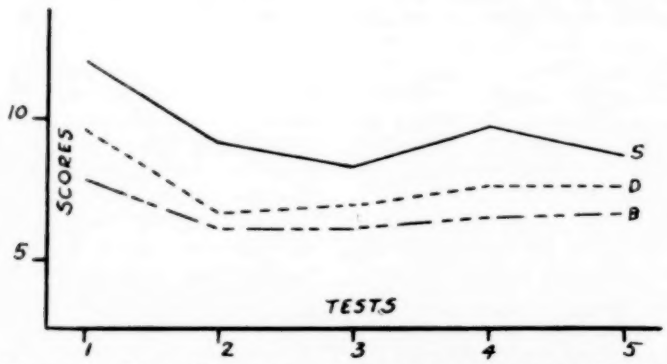



Figure 5.—The total average scores made on each test by the S, D, and B method of presenting terms—G2.

(2) The variations in retention for the different tests.—If learning by the conditioned reflex technique is unstable and soon lost, this condition will be apparent when the total average scores for each test are compared. These data are given in Table IX. The highest total average scores are registered on test 1, the test immediately following the presentation of terms. It will be noted that there is a considerable drop in the total average scores from test 1 to test 2 and that the statistical reliability of the differences between these averages is high. Retention is approximately the same for tests 2 and 3. For each of the five learning situations the total average scores are higher on test 4 than they were on test 3. The increase in retention on test 4 over tests 2 and 3 is shown by the rise in each of the five curves representing the five learning situations in Figure 6. This consistent increase in the total average scores on test 4 over test 3 is considerably higher than a chance increase, as is shown by the reliability of the differences between the average scores. Retention is slightly less efficient on test 5 (for interpolated pied type it is slightly better) than

## TABLE VIII

THE AVERAGE SCORES MADE BY THE SIMULTANEOUS, DELAYED, AND BACKWARD METHODS OF PRESENTING TERMS FOR EACH LEARNING SITUATION IN EACH TEST—G2

| | Test 1 | | | Test 2 | | | Test 3 | | | Test 4 | | | Test 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ave. | Diff. | Rel. | Ave. | Diff. | Rel. | Ave. | Diff. | Rel. | Ave. | Diff. | Rel. | Ave. | Diff. | Rel. |
| UNFILLED | | | | | | | | | | | | | | | |
| S | 15.18 | | | 12.54 | | | 11.91 | | | 13.09 | | | 11.00 | | |
| D | 14.65 | .53 | .23 | 10.90 | 1.64 | .67 | 11.27 | .64 | .24 | 11.73 | 1.36 | .53 | 11.44 | .44 | .18 |
| B | 13.82 | .83 | .33 | 11.00 | .10 | .04 | 11.36 | .09 | .04 | 11.45 | .28 | .11 | 10.44 | 1.00 | .49 |
| S | 15.18 | 1.36 | .72 | 12.54 | 1.54 | .73 | 11.91 | .55 | .23 | 13.09 | 1.64 | .73 | 11.00 | .56 | .29 |
| PIED TYPE | | | | | | | | | | | | | | | |
| S | 12.25 | | | 9.19 | | | 7.60 | | | 8.93 | | | 8.86 | | |
| D | 9.93 | 2.32 | 1.36 | 7.00 | 2.19 | 1.37 | 7.13 | .47 | .25 | 8.64 | .29 | .15 | 8.71 | .15 | .08 |
| B | 8.13 | 1.80 | 1.03 | 6.56 | .44 | .27 | 6.27 | .86 | .49 | 7.00 | 1.64 | .94 | 7.36 | 1.35 | .78 |
| S | 12.25 | 4.12 | 2.42 | 9.19 | 2.63 | 1.63 | 7.60 | 1.33 | .83 | 8.93 | 1.93 | 1.05 | 8.86 | 1.50 | .85 |
| ARITHMETIC | | | | | | | | | | | | | | | |
| S | 12.71 | | | 8.59 | | | 7.90 | | | 9.38 | | | 8.19 | | |
| D | 10.41 | 2.30 | 1.47 | 5.53 | 3.06 | 2.81 | 5.70 | 2.20 | 1.68 | 6.12 | 3.26 | 2.69 | 6.06 | 2.13 | 1.53 |
| B | 7.00 | 3.41 | 2.10 | 5.41 | .12 | .11 | 5.36 | .34 | .27 | 5.75 | .37 | .25 | 6.63 | .57 | .39 |
| S | 12.71 | 5.71 | 3.88 | 8.59 | 3.18 | 3.00 | 7.90 | 2.54 | 1.95 | 9.38 | 3.63 | 2.77 | 8.19 | 1.56 | 1.16 |
| PICTURES | | | | | | | | | | | | | | | |
| S | 10.63 | | | 8.21 | | | 7.68 | | | 9.13 | | | 8.38 | | |
| D | 7.83 | 2.80 | 1.82 | 6.21 | 2.00 | 1.71 | 6.86 | .82 | .57 | 7.04 | 2.09 | 1.57 | 7.00 | 1.38 | .94 |
| B | 5.12 | 2.71 | 1.87 | 3.75 | 2.46 | 2.36 | 3.14 | 3.72 | 2.77 | 3.48 | 3.56 | 3.36 | 4.19 | 2.81 | 2.18 |
| S | 10.63 | 5.51 | 4.34 | 8.21 | 4.46 | 4.25 | 7.68 | 4.54 | 4.41 | 9.13 | 5.65 | 5.54 | 8.38 | 4.19 | 3.84 |
| NAMES | | | | | | | | | | | | | | | |
| S | 9.40 | | | 7.20 | | | 6.10 | | | 7.25 | | | 6.60 | | |
| D | 5.45 | 3.95 | 4.25 | 3.50 | 3.70 | 3.59 | 3.60 | 2.50 | 2.55 | 3.70 | 3.55 | 3.09 | 4.25 | 2.35 | 2.16 |
| B | 4.85 | .60 | .54 | 3.60 | .10 | .13 | 4.05 | .45 | .54 | 3.85 | .15 | .17 | 3.80 | .45 | .48 |
| S | 9.40 | 4.55 | 4.30 | 7.20 | 3.60 | 3.30 | 6.10 | 2.05 | 1.95 | 7.25 | 3.40 | 2.86 | 6.60 | 2.80 | 2.64 |
| AVERAGE | | | | | | | | | | | | | | | |
| S | 12.03 | | | 9.15 | | | 8.24 | | | 9.56 | | | 8.61 | | |
| D | 9.67 | 2.36 | 2.95 | 6.63 | 2.52 | 3.55 | 6.91 | 1.33 | 1.75 | 7.45 | 2.11 | 2.78 | 7.49 | 1.12 | 1.50 |
| B | 7.78 | 1.89 | 2.28 | 6.06 | .57 | .80 | 6.03 | .88 | 1.19 | 6.31 | 1.14 | 1.54 | 6.48 | 1.10 | 1.38 |
| S | 12.03 | 4.25 | 5.90 | 9.15 | 3.09 | 4.70 | 8.24 | 2.21 | 3.20 | 9.56 | 3.25 | 4.58 | 8.61 | 2.13 | 3.18 |

## TABLE IX

THE TOTAL AVERAGE SCORES MADE ON EACH TEST IN EACH LEARNING SITUATION—G2

| | T1 | | T2 | | T3 | | T4 | | T5 | | T1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| UNFILLED | | | | | | | | | | | |
| Ave. | 43.65 | | 34.44 | | 34.54 | | 36.27 | | 32.88 | | 43.65 |
| Diff. | | 9.21 | | .10 | | 1.73 | | 3.39 | | 10.77 | |
| Rel. | | 1.79 | | .02 | | .26 | | .55 | | 1.94 | |
| PIED TYPE | | | | | | | | | | | |
| Ave. | 30.31 | | 22.75 | | 21.00 | | 24.57 | | 24.93 | | 30.21 |
| Diff. | | 7.56 | | 1.75 | | 3.57 | | .36 | | 5.28 | |
| Rel. | | 1.68 | | .38 | | .77 | | .04 | | 1.13 | |
| ARCHITECT | | | | | | | | | | | |
| Ave. | 30.12 | | 19.53 | | 18.96 | | 21.25 | | 20.88 | | 30.12 |
| Diff. | | 10.59 | | .57 | | 2.29 | | .37 | | 9.24 | |
| Rel. | | 2.32 | | .13 | | .67 | | .10 | | 2.58 | |
| PICTURES | | | | | | | | | | | |
| Ave. | 23.58 | | 18.17 | | 17.68 | | 19.65 | | 19.57 | | 23.58 |
| Diff. | | 5.41 | | .49 | | 1.97 | | .08 | | 4.01 | |
| Rel. | | 1.55 | | .16 | | .65 | | .03 | | 1.12 | |
| NAMES | | | | | | | | | | | |
| Ave. | 19.70 | | 14.30 | | 13.75 | | 14.80 | | 14.65 | | 19.70 |
| Diff. | | 5.40 | | .55 | | 1.05 | | .15 | | 5.05 | |
| Rel. | | 2.29 | | .19 | | .42 | | .06 | | 2.02 | |
| AVERAGES | | | | | | | | | | | |
| Ave. | 29.47 | | 21.82 | | 21.18 | | 23.31 | | 22.58 | | 29.45 |
| Diff. | | 7.65 | | .64 | | 2.13 | | .73 | | 6.87 | |
| Rel. | | 3.94 | | .35 | | 1.09 | | .38 | | 3.52 | |

on test 4, but the reliability of these differences between the total average scores is only slightly better than chance.

A comparison of the total average scores for each learning situation for each test (bottom of Table IX) shows the gain in retention on test 4 over tests 2, 3, and 5. The reliability of the difference between the total average scores made on tests 1 and 2 has complete statistical reliability. The reliability of the difference between the scores of tests 2 and 3 is low. The gain in total average score on test 4 over test 3 is sufficient to show a reliability greater than chance. The difference between the total average scores for tests 4 and 5 possesses only slightly greater than chance reliability.

The fact that retention was better on test 4, three weeks after learning, than it was on test 2 or 3, one and two weeks respectively after learning, may represent what Ballard (2) called "reminiscence". If these data do illustrate the phenomenon of reminiscence, they do not support the hypothesis that reminiscence is characteristic of the memory of children for poetry. The subjects in this study were not children and the learning materials were not poetry; and the experimental conditions were quite different from those of Ballard's.

These data show the importance of time in relation to efficiency of retention. If materials of the type learned in this study cannot be tested immediately after the learning period, then the opportune time for testing is three weeks after the learning period.

(3) Variations in retention in relation to the interpolated activities.—A comparison of the scores for each test of the filled learning situations with those of the unfilled (Table IX) shows that, although less is learned in the filled than in the unfilled learning situation, the amount retained for each test in the filled learning situations varies little from that retained from what was learned in the unfilled. The constancy of retention, regardless of the nature of the interpolated activities, may be readily seen by examining the curves in Figure 6, the curves plotted from the data in Table IX.

These curves show that learning in relation to the interpolated activities is most efficient in the following order: unfilled, pied type, arithmetic, pictures, and names. The distance between these curves shows the relative efficiency of learning and retention for each test. It will be noted that the curve representing the total average scores made on the unfilled learning situation is well above that of the pied type; that the curves for pied type, arithmetic, and pictures are relatively close together; and that the curve for names is well below that of pictures. These curves, as well as the data in Table IX, show that retention is constant regardless of whether the intervals between terms are filled or un-

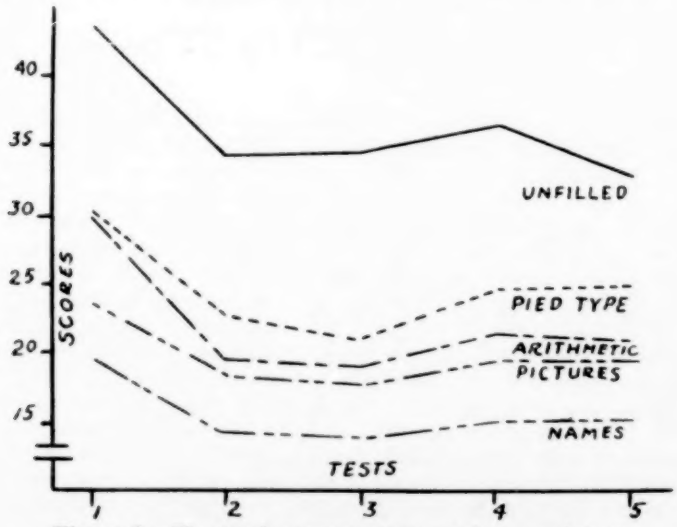



Figure 6.—The total average scores made on each test in each learning situation.

TABLE X

A COMPARISON OF THE TOTAL AVERAGE SCORES MADE BY THE IMMEDIATE AND RE-TEST GROUPS FOR EACH LEARNING SITUATION

| | Unfilled | | Pied type | | Names | | Pictures | | Arithmetic | |
|---|---|---|---|---|---|---|---|---|---|---|
| | G1 | G2 | G1 | G2 | G1 | G2 | G1 | G2 | G1 | G2 |
| Ave. | 45.56 | 43.65 | 38.71 | 19.70 | 37.50 | 19.70 | 33.64 | 23.58 | 21.81 | 30.12 |
| Diff. | 1.91 | | 17.80 | | 17.80 | | 10.06 | | 8.31 | |
| Rel. | .42 | | 4.62 | | 4.52 | | 2.56 | | 2.52 | |

filled. Neither does the nature of the interpolated activities affect the efficiency of retention.

(4) A comparison of the efficiency of learning of the re-test group with that of the immediate test group.—Because of differences in number of tests and length of time intervals between terms, possibilities for comparison between the two groups are limited. The immediate test group had only the one test, immediately following the learning situation; and the efficiency of learning of the re-test group was studied at only one time interval, the 30-second. Therefore, comparisons between the two are on the basis of the 30-second interval in the test immediately following the learning situation. The comparisons are further limited to a consideration of the total average scores (the sum of the average scores on the simultaneous, delayed, and backward method) in each learning situation.

The efficiency of learning for the different learning situations for the immediate test group ($G_1$) was interpolated rest first, followed by names, pied type, pictures, and arithmetic. For the re-test group ($G_2$) learning was most efficient for interpolated rest, followed by pied type, arithmetic, pictures, and names. The total average scores of $G_1$ and $G_2$ are given in Table X. The $G_1$ scores are higher in all learning situations except interpolated arithmetic. The reliability of the difference between the total average scores in the unfilled learning situation is low, but in the other four learning situations the reliabilities of the differences are high.

The curves in Figure 7, plotted from the data in Table X, are close together in interpolated rest, showing $G_1$ slightly superior to $G_2$.

The superiority of learning of $G_1$ and $G_2$ increases through pied type to names, where

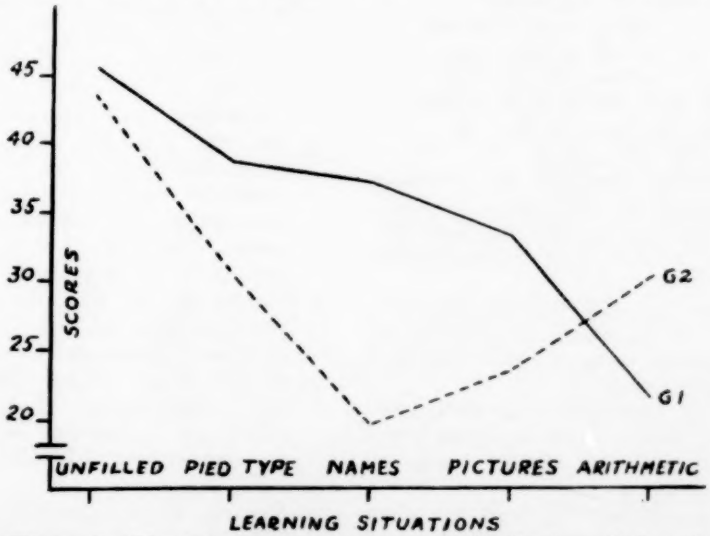



Figure 7.—The total average scores made by the immediate and re-test groups for each learning situation.

the distance between the two curves is greatest. Differences in extent of learning between the two groups is reduced in pictures; and for arithmetic the curves have crossed, $G_2$ being superior to $G_1$.

The marked superiority of learning of $G_1$ over $G_2$ in pied type, names, and pictures, and approximately as great a superiority of $G_2$ over $G_1$ in arithmetic, is probably additional evidence that reduction in learning resulting from interpolated activities is dependent upon the degree of concentration of the subjects in the activities and not upon the nature of the activities.

## Summary

To investigate the efficiency of learning a less specialized type of behavior by the conditioned reflex technique, 608 subjects, 520 of whom were tested only immediately following the learning period and 88 of whom in addition to the immediate test were given four additional tests distributed over a period of five weeks, were tested on their ability to learn and remember the names of 30 pictures. Five different time intervals, 20, 30, 40, 50, and 60 seconds, between the presentation of terms were studied; and the effect on learning efficiency of interpolated rest, working arithmetic problems, checking letters in pied type, memorizing names, and learning the names of other pictures, at each time interval, was studied. Some of the more important results of the study are as follows:

1. There are as many variations from the laws established for the conditioning of the true reflexes as there are similarities; therefore, it would seem that the type of learning studied in this investigation does not follow the laws of conditioning of the true reflex.

2. Of the three methods of presenting terms (simultaneous, delayed, and backward), simultaneous provides the most efficient learning; the backward method is least efficient.

3. Time intervals of different lengths produced changes in the efficiency of learning by the simultaneous, delayed, and backward methods of presenting terms.

4. Increasing the length of the time interval (unfilled) between terms increased learning efficiency, the most efficient being the 40-second interval.

5. Learning efficiency was reduced when the interval between the presentation of terms was filled with activities.

6. Interpolated activities between terms reduced learning efficiency in proportion to the amount of attention (concentration) demanded. The similarity or lack of similarity of the interpolated activities to the major learning materials seemed not to be important.

7. Changes in the relationship of the efficiency of the simultaneous, delayed, and backward methods of presenting terms and variations in learning efficiency at the different time intervals resulted from the interpolated activities.

8. Retention was not affected by the method (simultaneous, delayed, and backward) of presenting terms.

9. Retention was not affected by the interpolated activities.

10. Learning was stable with no tendency for the new responses to fade away.

## Bibliography

1. Allport, F. H. *Social Psychology*. Boston: Houghton Mifflin Co., 1924.
2. Ballard, P. B. *Obliviscence and Reminiscence*. Journal of Psychology Monograph, Supplement, Vol. I, No. 2, 1913. Pp. 82.
3. Bergstrom, G. A. "Effect of Changes in the Time Variables in Memorizing," *American Journal of Psychology*, XVIII (April, 1907), 206–38.
4. Book, William F. *Economy and Technique of Learning*, Ch. XI. Boston: D. C. Heath and Co., 1932.
5. Cason, Hulsey. "The Conditioned Pupillary Reaction," *Journal of Experimental Psychology*, V (April, 1922), 108–46.
6. DeCamp, J. Edgar. *A Study of Retroactive Inhibition*. Psychological Monograph, No. 84, August, 1915.
7. Froeberg, S. "Simultaneous Versus Successive Association," *Psychological Review*, XXV (March, 1918), 156–63.
8. Garrett, H. E. *Statistics in Psychology and Education*, pp. 128–34. New York: Longmans, Green and Co., 1926.
9. Heine, Rosa. "Uber Wiedererkennen und ruckwirkende Hemmung," *Zeitschrift fur Psychologie und Physiologie der sinnesorgane*, LXVIII (1914), 161–236.
10. Hilgard, E. R., and Campbell, A. A. "The Course of Acquisition and Retention of Conditioned Eyelid Responses in

Man," *Journal of Experimental Psychology*, XIX (1936), 227f.

11. Jones, Mary Cover. "The Elimination of Children's Fears," *Journal of Experimental Psychology*, VII (October, 1924), 382–90.
12. Moss, F. A., Editor. *Comparative Psychology*, pp. 367–402. New York: Prentice–Hall, 1934.
13. Müller, G. E., and Pilzeker, A. "Experimentelle Beitrage Zur Lehre Vom Gedachtniss," *Zeitschrift fur Psychologie und Physiologie der sinnesorgane*, Erganzungheft I (1900), 1–300.
14. Pavlov, Ivan P. *Lectures on Conditioned Reflexes: Twenty-five Years of Objective Study of the Higher Nervous Activity (Behavior) of Animals*. Translated by W. H. Grant. New York: International Publishers, 1928.
15. Robinson, E. S. "The Similarity Factor in Retroaction," *American Journal of Psychology*, (Washburne Commemorative volume), XXXIX (December, 1927), 297–312.
16. Smith, Stevenson, and Guthrie, E. R. *General Psychology in Terms of Behavior*. New York: D. Appleton Co., 1921.
17. Snoddy, G. S. *Evidence for Two Opposed Processes in Mental Growth*. Lancaster, Pa.: The Science Press Printing Co., 1935.
18. Thorndike, Edward L. *Human Learning*. New York: The Century Co., 1931.
19. Tolman, Edward C. *Purposive Behavior in Animals and Men*. New York: The Century Co., 1932. Pp. 368.
20. Watson, John B. *Behaviorism* (Revised Edition). New York: W. W. Norton and Co., 1930.
21. Watson, John B., and Rayner, Rosalie. "Conditioned Emotional Reactions," *Journal of Experimental Psychology*, III (February, 1920), 1–14.
22. Wolfle, Helen Morrill. "Time Factors in Conditioning of Finger-Withdrawal," *Journal of General Psychology*, IV (December, 1930), 372–8.

# VARIATIONS OF NORMAL CHILDREN*

HUMBERT G. STOUT
*Nebraska State Teachers College*
*Kearney*

The purpose of this study was to determine as nearly as possible to what extent ten-year-old children of normal intelligence differ among themselves and within themselves in such traits as physical health, socio-economic status, mechanical ability, musical ability, school achievement, and behavior tendencies. The problem was first suggested because of the apparent lack of sufficient accurate knowledge about this kind of child. Because the writer was of the opinion that educational workers could use more data in this area, he set out to investigate along a line that appeared to have been somewhat neglected.

It is very important for the organization of the schools that superintendents, principals, and supervisors know whether or not the large masses of pupils can be grouped for effective work. If, as is contended, there is no average child, if every pupil is a law unto himself, then there should be more schools like Winnetka. If, on the other hand, some satisfactory principle of homogeneous grouping can be found, then much time, energy, and expense may be saved by such grouping.

An examination of 1,312 titles listed in the bibliographies of the *Review of Educational Research* showed the following numbers of articles devoted to children of the three intellectual levels:

20 articles dealing with superior children
38 articles dealing with dull children
11 articles dealing with normal children.

A canvass of the *Psychological Abstracts* from the beginning of the publication in 1927, up to and including January, 1937, disclosed the following numbers of articles dealing with some phase of variations of other traits in relation to intelligence:

71 articles dealing with superior children
99 articles dealing with dull children
27 articles dealing with normal children.

The profiles printed in Paul R. Mort's book *The Individual Pupil*[1] showed these groups:

24 pupils with intelligence quotients above 110
23 pupils with intelligence quotients below 90
20 pupils with intelligence quotients between 90 and 110.

Of 1,543 titles of "Doctors' Theses Under Way in Education" listed in the *Journal of Educational Research* for the years 1934–1937, only two studies devoted to the normal child are listed—one of these studies being the present one. A careful examination of the studies found in all the sources investigated discloses the fact that in only three of them has "normality" been based primarily upon intelligence quotients between 90 and 110. None of the studies bears directly upon the differences of ten-year-old children of normal intelligence. From the above consideration it seems clear that:

1. In the past, superior children and dull children have proved relatively more interesting than normal children for those making scientific studies, since almost two and one half times as many studies are reported referring to superior children, and more than three and one half times as many studies of dull children as those referring to normal children.

2. Few studies of strictly normal children are reported. Being just "normal," they have evidently attracted little attention.

3. There has not been any increased interest during the past four years in the normal child. It appears, therefore, that people working in the field of research in education have neglected the child of normal intelligence; and theoretically, one half of the children of the country are in this neglected group. In short, there is a broad field and an apparent need for further study of children of normal intelligence.

For the purpose of this study, ten-year-old children whose intelligence quotients lay be-

* This article is a condensation of a doctor's dissertation (1937) by the author, a complete copy of which is on file in the office of Dr. D. A. Worcester, Chairman of the Department of Educational Psychology and Measurements, at the University of Nebraska. The original manuscript contains thirty-three tables and 129 pages of profiles and graphs. Any one interested in examining the data in detail may consult the manuscript or write the author.

[1] Paul R. Mort, *The Individual Pupil*. New York: The American Book Company, 1928.

tween 90 and 110 were selected. The Haggerty Intelligence Examination, Delta 2, for grades 3–9 and the Otis Self-Administering Tests of Mental Ability for grades 4–8 were given to all ten-year-old pupils in the Public Schools of Kearney, Nebraska, and in the A. O. Thomas School—the latter being the training school of the Nebraska State Teachers College at Kearney. These pupils had previously been checked in the Bureau of Vital Statistics at Lincoln, Nebraska, for age. The Stanford–Binet Individual Examination (1916) was then given to twenty of those pupils whose I.Q.'s were shown by the two group tests to be very far apart, or whose I.Q.'s the group tests showed to be near 90 or 110. Only those were chosen for the study whose mean I.Q.'s were between 90 and 110. This procedure made available a group of fifty-eight ten-year-old pupils of normal I.Q. (median = 98.8), consisting of thirty boys and twenty-eight girls who were used as the subjects for this study.

## Health and Physical Condition

The "Health Record" and the "Personal History—Contagious Diseases" record for each of these pupils were examined. These data were obtained from the files of the trained nurse of the Kearney City Schools and from the office of the Director of Teacher Training at the State Teachers College. The health records included data relative to height, weight, and defects in eyes, ears, teeth, throat, nose, glands, skin, posture, feet, and nutrition. Table I shows in tabular form portions of the data relative to these traits. The totals at the bottom of the table indicate the general trends of the different items included in the table. One notes that there are eighteen cases of severe defects as shown by the mark xxx. For example, one pupil, AP, has four severe defects,—in throat, nose, posture, and nutrition; and three slight defects in teeth, skin, and feet. Pupil CA is the only one with no defects marked against him. He has, however, had his tonsils removed.

The heights and weights are shown in the second and third columns of the table. The heights vary from forty-six and one half inches to fifty-nine inches, and the weights from forty-five pounds to eighty-nine pounds. Pupil AA is fifty-four inches tall, weighs eighty-two pounds, and the standard for his height at age ten years is seventy pounds. He is, therefore, twelve pounds overweight. Pupil AC is forty-nine and one half inches tall, weighs fifty-five pounds, and is at the correct weight for his age and height. As would be expected, there is a greater variation in weight than in height. A range of twelve and one half inches in a group of ten-year-olds whose mean height is fifty-three inches, and a variation of forty-four pounds in the same group whose mean weight is sixty-six pounds, indicate that the members

TABLE I

HEALTH RECORDS OF SIX AND SUMMARY FOR FIFTY-EIGHT TEN-YEAR-OLD CHILDREN OF NORMAL INTELLIGENCE

| Pupil | Ht. | Wt. | Eyes | Ears | Teeth | Throat | Nose | Glands | Skin | Posture | Feet | Nutrition |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | 54 | 82–70* | ✓ | ✓ | ✓ | oo | oo | ✓ | ✓ | x | ✓ | ✓ |
| AD | 49 | 50–56 | x | ✓ | x | oo | ✓ | xx | ✓ | x | ✓ | ✓ |
| AP | 57 | 81–82 | ✓ | ✓ | x | xxx | xxx | ✓ | x | xxx | x | xxx |
| AQ | 54 | 80–70 | ✓ | ✓ | x | xxx | xxx | ✓ | ✓ | x | ✓ | x |
| CA | 46½ | 47–50 | ✓ | ✓ | ✓ | oo | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| CB | 53 | 84–67 | x | ✓ | xx | xx | x | x | ✓ | x | ✓ | ✓ |
| Totals | | | | | | | | | | | | |
| xxx | | | | | 3 | 6 | 4 | | | 3 | | 2 |
| xx | | | 3 | | 10 | 17 | 5 | 3 | | 10 | | 1 |
| x | | | 15 | | 19 | 12 | 13 | 7 | 4 | 26 | 2 | 9 |
| ✓ | | | 35 | 48 | 22 | 4 | 25 | 38 | 50 | 9 | 46 | 36 |
| oo | | | | | | 15 | 7 | | | | | |
| No record | | | 5 | 10 | 4 | 4 | 4 | 10 | 4 | 10 | 10 | 10 |
| M | 52.66 | 66.09 | | | | | | | | | | |
| Md | 52.75 | 66.25 | | | | | | | | | | |
| Q | 1.95 | 8.44 | | | | | | | | | | |
| S. D. | 2.83 | 10.99 | | | | | | | | | | |
| V | 5.37 | 16.61 | | | | | | | | | | |
| Range | 12.5 | 44.00 | | | | | | | | | | |

Code: ✓—normal, x—slight defect, xx—moderate defect, xxx—severe defect, 00—removed.
* The last number is the standard weight for the height; e. g., AA is 54 inches tall, weighs 82 pounds, and the standard weight for 54 inches at age 10 is 70 pounds.

TABLE II

PERSONAL HISTORY—CONTAGIOUS DISEASES AND VACCINATION OF SIX, AND TOTALS FOR FIFTY-EIGHT TEN-YEAR-OLD CHILDREN OF NORMAL INTELLIGENCE*

| Pupil | Chicken Pox | Measles | Whooping Cough | Mumps | Scabies | Scarlet Fever | Small pox | Diphtheria | Pneumonia | Eczema | Small pox vac. | Diphtheria vac. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | ✓ | ✓ | ✓ | | | | | | | | ✓ | ✓ |
| AC | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | | ✓ |
| | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| BA | ✓ | ✓ | ✓ | | | | | | | | | ✓ |
| BB | ✓ | | ✓ | ✓ | | | | | | | | ✓ |
| | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| CA | | | | | | | | | | | | ✓ |
| CB | ✓ | ✓ | ✓ | ✓ | | | | | | ✓ | | ✓ |
| Total | 34 | 26 | 29 | 23 | 1 | 4 | 7 | 1 | 7 | 3 | 11 | 43 |

* It is of interest to note that there have been only eleven of the children vaccinated for small-pox, and there are seven cases of the disease; whereas, there have been forty-three vaccinated for diphtheria with only one case of the disease.

of this group are far from uniform in height and weight.

Table II shows the pupils' personal history relative to contagious diseases, and vaccination for small-pox and diphtheria. Over one half of the pupils have had chicken-pox and whooping cough; nearly one half have had measles and mumps; one pupil has had scabies; four have had scarlet fever; and seven have had small-pox. There have been forty-three cases of vaccination for diphtheria and eleven for small-pox. AP, the most defective (physically) boy, has had no diseases and has not been vaccinated, whereas the most nearly normal (physically) boy has had no diseases and has been vaccinated for diphtheria.

Consideration of the data gathered relative to the health and physical condition of these fifty-eight ten-year-old children of normal intelligence discloses that in this group: (1) There is approximately three times as much variation in weight as in height; (2) there is a positive correlation of $.822 \pm .029$ between height and weight; (3) there is no significant relationship between either height or weight and number of physical defects; and (4) the fact that a child is normal in intelligence gives little or no clue as to what his health or physical condition may be.

## SOCIO-ECONOMIC STATUS

The pupils were asked to fill out the Sims Score Card for Socio-Economic Status,[2] since this method seemed more reliable and also more valid than would have been any attempt to evaluate their home conditions by means of personal visits to the various homes. The pupils were taken in small groups, and each question was read to them and explained. They were permitted to ask questions freely in order that they might fully understand just what was wanted. The object was to secure as accurate information as possible from the children themselves. It is not thought that any of them tried to misrepresent the facts. They all seemed sincere and honest in their endeavor to answer the questions as correctly as possible. While, in the opinion of the writer, the results obtained by the use of this score card are influenced somewhat by subjective elements, yet the method is far superior to any other instrument thus far devised.

The twenty-three questions on each pupil's score card have to do with the occupation, education, clubs, etc., of his parents; the size, the furnishings, and the conveniences of his home; how and where he spends his vacations; and the frequency and type of amusements in which he indulges. The twenty-three questions carry a possible total of eighty-two credit points. The total score is obtained by taking the quotient of the total credit points divided by the number of questions answered (usually twenty-three), carried to one decimal place and the decimal point dropped. Thus if the total credit is 65 the score is $(65 \div 23 = 2.8)$ 28. Sims finds that the coefficient of reliability for this score card is .91, and that measured by the best obtainable standards the score card is valid.

Samples of the data collected for these pupils are shown in Table III. Three pupils did not fill out the score card; two answered a total of twenty-one questions; fifteen answered twenty-two questions each; and each of the remaining thirty-eight answered the entire twenty-three questions. The highest score for any one pupil is 32 and the lowest is 5.

[2] Verner Martin Sims, *The Measurement of Socio-Economic Status*, Bloomington, Illinois: Public School Publishing Company, 1928.

TABLE III

STANDINGS OF SIX AND SUMMARY FOR FIFTY-EIGHT PUPILS OF NORMAL INTELLIGENCE IN SOCIO-ECONOMIC STATUS

| Pupil | Questions Answered | Total Credits | Score |
|---|---|---|---|
| BZ | 22 | 70 | 32 |
| AX | 23 | 68 | 30 |
| ... | ... | ... | ... |
| AD | 22 | 33 | 15 |
| CC | 23 | 35 | 15 |
| ... | ... | ... | ... |
| BQ | 22 | 14 | 6 |
| AI | 23 | 11 | 5 |
| Mean | | 37.32 | 16.42 |
| Median | | 35.00 | 15.57 |
| Q. | | 10.14 | 4.53 |
| S.D. | | 13.76 | 6.12 |
| V. | | 36.86 | 37.29 |
| Range | | 59.00 | 27.00 |

Thus we have a range, in these scores, of 27 points and a mean score of 16.42; i.e., the range is 1.6 times as great as the mean score. Sims warns against attaching too much significance to a comparison of local data with norms furnished. However, these scores, ranging from 5 to 32, with a quartile deviation of 4.53, and a standard deviation of 6.12, certainly show a wide variation in socio-economic status. We may sum up the findings from the data gathered relative to the socio-economic status of these fifty-eight normal pupils as follows: (1) They differ widely among themselves; (2) they vary greatly within themselves in the different elements of wealth and social standing; and (3) there is a slight positive relationship ($r = .286 \pm .086$) between socio-economic status and freedom from physical defects.

## GENERAL MECHANICAL ABILITY

The Stenquist Assembling Tests of General Mechanical Ability[3] were given to the pupils individually. The publishers of this series of tests recommend that they be given as a group test to save time. The writer found, however, that the price of the series made group testing prohibitive. The pupils were called into the office one at a time and given the tests. Otherwise, the directions were followed rigidly, and thus the conditions were kept uniform. Norms for ten-year-old pupils were not available, and hence the norms for nine-year-old boys were used. The Stenquist Assembling Test Series III is designed especially for boys, but was given to the entire group. It consists of a plain bolt and nut, a bolt and wing nut, a plain hinge, a key and ring, a turn buckle, a drawer pull, a trunk caster, a plain push button, a belt and buckle, and a nail clip. These ten articles were taken apart, and the parts of each put into a separate compartment in a wooden case. The task of the pupil was to assemble the articles in order, being allowed thirty minutes time for the entire test.

This test was given to all of the boys (thirty) and twenty-six of the girls. Table IV is an abridged form of the original table. The raw scores are used rather than the T-scores because the only available table of standardized norms for Series III was apparently standardized on the raw scores. At the time the *Manual of Directions* was published, 1922, Series III had not been standardized, but it has since been standardized by Edna Willis McElwee.[4] McElwee evidently did not use the T-scores, since the scores in her table of norms range from 18 to 105. The possible range of raw scores is from 0 to 120, whereas the possible range of T-scores is from 49 to 82. Each of the first eight models used in the Stenquist test carries a possible credit of ten points, while models I and J carry twenty points each. Consequently, there is a possible total of 120 points.

The data indicate that: (1) The variation in mechanical ability among boys and girls of normal intelligence is very great; (2) girls vary slightly more than boys; (3) there is no significant relationship between mechanical ability and freedom from physical defects ($r = -0.175 \pm .092$); and (4) there is no significant relationship between mechanical ability and socio-economic status ($r = .111 \pm .090$).

## MUSICAL ABILITY

To measure the pupils in musical ability, the Kwalwasser-Dykema Music Tests[5] were used. For each of the ten tests, a phonograph record was played, and the pupils were asked to indicate by certain letters whether tones, pitches, rhythms, melodies, etc., were alike or different. The scores were translated into percentile rankings for comparison with standard norms. There is some ques-

[3] John L. Stenquist, *Stenquist Assembling Test of General Mechanical Ability*. Chicago, Illinois: C. H. Stoelting Co., 1922.

[4] Edna Willis McElwee, "Standardization of the Stenquist Assembling Test, Series III," *Journal of Educational Psychology*, XXIII (September, 1932), 451–54.

[5] K-D Music Tests, Published by Carl Fischer, Inc., Cooper Square, New York, 1930.

TABLE IV

SCORES FOR SIX AND SUMMARY FOR FIFTY-SIX PUPILS OF NORMAL INTELLIGENCE IN THE STENQUIST ASSEMBLING TESTS

| Pupil | Model A | Model B | Model C | Model D | Model E | Model F | Model G | Model H | Model I | Model J | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BL | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 1 | 20 | 20 | 111 |
| AK | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 4 | 10 | 20 | 104 |
| ... | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| CF | 10 | 10 | 2 | 0 | 10 | 10 | 10 | 0 | 0 | 10 | 62 |
| CC | 10 | 10 | 10 | 0 | 10 | 6 | 10 | 0 | 5 | 0 | 61 |
| ... | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| AN | 10 | 10 | 2 | 2 | 10 | 1 | 0 | 0 | 0 | 0 | 35 |
| AC | 10 | 10 | 0 | 0 | 2 | 6 | 0 | 0 | 0 | 0 | 28 |

| | Boys | Girls | Total |
|---|---|---|---|
| Mean | 68 | 53.7 | 61.3 |
| Median | 66 | 52.5 | 61.1 |
| Q | 11.5 | 10.9 | 10.4 |
| S. D. | 16.7 | 13.8 | 16.9 |
| V | 24.5 | 25.6 | 27.6 |
| Range | | | 83 |

tion as to how much the tests measure nature, and how much they measure nurture; but so far as we are able to determine, no one doubts that they distinguish fairly accurately the pupil's musical ability. The ten tests in the series include tonal memory, quality discrimination, intensity discrimination, tonal movement, time discrimination, rhythm discrimination, pitch discrimination, melodic taste, pitch imagery, and rhythm imagery. The scores are presented in Table V. The scores of the ten tests are not equated; hence, they cannot be compared directly. Moreover, the means vary from 12.5 to 23.5, so the standard deviations cannot be compared directly. Therefore, it seems that comparisons of the variations in the different parts of the music tests are best made by means of the Pearson formula for calculating the coefficient of variation. This formula[6] is $V = \frac{S.\ D.\ \times\ 100}{Average}$ and simply amounts to finding the ratio between the S.D. and the mean, and stating it in terms of per cent.

If we examine the different V's in this series of music tests, we shall find that the lowest is for pitch imagery and is 14.6. To state this a little differently, at least one third of the scores in each of these tests depart from the mean for the test by more than 14.6% of that mean. If we go, now, to the other extreme, we shall find in the case of tonal movement that one third of the scores are away from the mean by more than 28.6% of the mean. When we total all the scores, we have a mean that is ten times as large as the means of the individual tests. However, the high scores in one test offset the low scores in another test, and we have a V of 9.1. The V's of these ten tests indicate that there is more variability within the tests themselves than the total V shows. In the light of these considerations we must conclude that the variability of this group of pupils is so great as to warrant individual attention to each pupil.

Furthermore, the $r$ between total music and physical defects was found to be zero; between music and socio-economic status it was $.429 \pm .084$; and between music and mechanical ability $.116 \pm .092$. These correlations indicate that there is no significant relationship between musical ability and any one of these other traits with the exception of socio-economic status. Although the $r$ between musical ability and socio-economic status is low and would be of no predictive value, it probably does carry some significance. If musical ability is partly due to environment, the better environment that goes with a better social and economic condition would doubtless furnish a better opportunity for the pupil to advance in musical performance and appreciation.

From the foregoing considerations we may say that the data collected relative to musical ability indicate that: (1) Individual differences in general musical ability are very

[6] There has recently developed considerable objection to the use of the Pearson formula, largely because of the uncertainty of the zero point. However, no better formula has been offered and it is used here for want of a better.

TABLE V

SCORES OF SIX PUPILS AND SUMMARY FOR FIFTY-EIGHT CHILDREN OF NORMAL INTELLIGENCE IN THE KWALWASSER–DYKEMA MUSIC TEST

| Pupil | Tonal Mem. | Quality | Intensity | Tonal Move. | Time | Rhythm | Pitch | Taste | Pitch Imag. | Rhythm Imag. | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BW | 20 | 25 | 25 | 25 | 12 | 22 | 29 | 14 | 14 | 17 | 203 |
| BZ | 19 | 20 | 25 | 27 | 19 | 19 | 27 | 13 | 14 | 19 | 202 |
| BP | 14 | 24 | 24 | 12 | 11 | 13 | 24 | 12 | 14 | 18 | 166 |
| CF | 14 | 17 | 21 | 20 | 13 | 18 | 20 | 13 | 12 | 18 | 166 |
| BV | 14 | 19 | 8 | 8 | 19 | 17 | 22 | 7 | 13 | 12 | 139 |
| AU | 17 | 8 | 9 | 20 | 11 | 13 | 16 | 12 | 12 | 19 | 137 |
| Mean | 15.4 | 19.6 | 21.0 | 16.8 | 14.3 | 16.7 | 23.5 | 12.5 | 14.3 | 15.8 | 165.7 |
| Md. | 15.1 | 19.5 | 21.6 | 16.9 | 13.4 | 16.2 | 23.8 | 12.8 | 14.4 | 15.6 | 164.3 |
| Q. | 2.1 | 1.7 | 1.9 | 3.5 | 1.9 | 1.9 | 2.4 | 1.7 | 1.5 | 2.1 | 9.2 |
| S. D. | 2.8 | 2.9 | 3.5 | 4.8 | 2.9 | 2.6 | 3.9 | 2.9 | 2.1 | 2.7 | 15.0 |
| V. | 18.0 | 14.9 | 16.6 | 28.6 | 20.6 | 15.3 | 16.7 | 23.1 | 14.6 | 17.1 | 9.1 |

great; (2) the same pupil is likely to show great individual differences in various phases of musical ability; (3) pupils tend to show greater individual differences in tonal movement than in any of the other traits tested; (4) there is a slight positive relationship between musical ability and socio-economic status, which fact probably has some significance; and (5) the best, and probably the only, safe way to determine a pupil's musical ability is to test it directly.

## ACHIEVEMENT IN THE COMMON BRANCHES

The New Stanford Achievement Test[7] was used in this investigation because of its widespread acceptance; because of its carefully prepared norms; and because of the fact that the scores on the ten tests are equated in value, making it possible to compare a pupil's scores in the various subjects by comparing them directly. The norms are so arranged that they show in parallel columns the standards for chronological age, for educational age, and for school grade. The comparisons in this particular study were made in terms of the school grade norms. In addition to the ten common subjects that are measured by the New Stanford Achievement Test, the pupil's rate and quality in handwriting were evaluated in terms of the American Handwriting Scale. In this manner scores were obtained for each pupil in paragraph meaning, word meaning, dictation, language usage, literature, history and civics, geography, physiology and hygiene, arithmetic reasoning, arithmetic computation, and rate and quality of handwriting. The data gathered from the results of this test were arranged in tabular form in Table VI which shows the highest, lowest, and middle pupils, together with the usual statistical summary.

If we examine the summary at the close of the table, we shall find some interesting results. Since the test scores are equated, we can compare directly the means in the differ-

[7] Published by the World Book Company, Yonkers-on-Hudson, New York; and Chicago, 1931.

TABLE VI

SCORES OF SIX PUPILS AND SUMMARY FOR FIFTY-EIGHT CHILDREN OF NORMAL INTELLIGENCE IN THE NEW STANFORD ACHIEVEMENT TESTS

| Pupil | Par. Mean. | Word Mean. | Total Read. | Diet. | Lang. Usage | Lit. | Hist. and Civ. | Geog. | Phys. and Hyg. | Arith. Reas. | Arith. Comp. | Total Arith. | Total Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AH | 78 | 62 | 70 | 58 | 62 | 86 | 82 | 71 | 76 | 80 | 71 | 76 | 72.5 |
| AR | 82 | 75 | 78.5 | 78 | 62 | 55 | 76 | 84 | 70 | 74 | 67 | 70.5 | 72.3 |
| BB | 69 | 61 | 65 | 43 | 44 | 64 | 51 | 67 | 65 | 43 | 60 | 56.5 | 58.7 |
| BU | 67 | 60 | 63.5 | 67 | 80 | 52 | 20 | 48 | 64 | 64 | 65 | 64.5 | 58.7 |
| BK | 58 | 52 | 55 | 44 | 42 | 34 | 20 | 34 | 24 | 34 | 56 | 45 | 39.8 |
| BH | 27 | 45 | 36 | 51 | 20 | 20 | 20 | 50 | 24 | 34 | 42 | 38 | 33.2 |
| Mean | 61.6 | 61.4 | 61.2 | 56.4 | 57.1 | 51.4 | 51.3 | 58.4 | 57.1 | 59.1 | 60.5 | 59.8 | 57.1 |
| Md. | 64.2 | 62.3 | 62.5 | 56.9 | 60.8 | 55.2 | 56.7 | 59.4 | 62.1 | 60.2 | 62.5 | 61.1 | 58.4 |
| Q. | 7.9 | 5.4 | 5.7 | 9.8 | 13.7 | 17.2 | 18.2 | 6.5 | 10.8 | 9.8 | 4.4 | 6.8 | 8.1 |
| S. D. | 12.5 | 9.0 | 9.8 | 13.2 | 20.0 | 19.7 | 20.1 | 11.8 | 17.6 | 13.5 | 9.5 | 9.9 | 9.9 |
| V. | 20.2 | 14.7 | 16.0 | 23.4 | 35.0 | 38.3 | 39.1 | 20.2 | 30.8 | 22.9 | 15.7 | 16.7 | 17.3 |

ent columns and arrive at the "group" differences. The lowest mean is 51.36 (history and civics), and the highest is 61.62 (paragraph meaning), while the mean of the total scores is 57.15. Accordingly, we find a range of 10.26 in the mean scores of the various tests. This amounts to 17.9% of the mean of the total scores. Because of this variation in the means of the diverse tests, we again resort to the use of the V as a means of comparing the dispersions of the ten tests. We find that the lowest V (14.76) is in word meaning, and that the highest (39.15) is in history and civics. In other words, in each of the ten tests one third of the group deviate from the mean of the group by from 15% to 39% of the amount of the mean. The range, moreover, in total scores is 39.3, and the highest score is two and one fifth times as great as the lowest score.

The following correlations between the total achievement scores and the other traits thus far studied are all negligible.

| | |
|---|---|
| Achievement—Height | .077 ± .091 |
| Achievement—Weight | .144 ± .089 |
| Achievement—Physical defects | .008 ± .093 |
| Achievement—Socio-economic status | .146 ± .089 |
| Achievement—Mechanical ability | .177 ± .088 |
| Achievement—Musical ability | .355 ± .083 |

In summary, the data relative to the achievement of these children of normal intelligence indicate that: (1) The variation of the group among the ten equated tests is considerable; (2) the individual differences of the pupils within themselves are great; (3) the individual differences of these pupils in their rankings in the different achievement tests are great with no general trend—except the tendency to vary; and (4) the relationships among the scores on the different achievement tests are positive and higher than between achievement scores and scores on any of the other traits studied, but are conspicuously low even among themselves.

## Behavior Tendencies

It was recognized from the start that the measurement of behavior is not as accurate as we should like. Yet, by use of the Haggerty-Olson-Wickman Rating Schedules,[8] ratings were secured that are valuable in estimating the conduct of these pupils. The schedules are two in number—A and B. Schedule A is designed to estimate the pupil's behavior in terms of overt acts performed; whereas, Schedule B is intended to estimate his tendencies rather than overt acts. Schedule B is divided into four divisions which consist of graphic ratings in intellectual, physical, social, and emotional traits. The scores are recorded in terms of totals for Schedule A, and for Schedule B, and for each of the four divisions of Schedule B. Schedule B is considered, by the authors of the instrument, as the more reliable of the two. The actual ratings of these fifty-eight pupils were made by their classroom teachers, and the data were assembled in Table VII.

In spite of its obvious subjectivity, Schedule B has, according to its authors, "a reliability of .92 as obtained from the correlation between halves of the scale." The authors further state that "ratings of the same pupils, when made by different judges under somewhat different conditions, commonly correlated about .60 for elementary school children." The originators of the schedules also warn that there is a "tendency to emphasize behavior of an aggressive type, and to miss certain non-aggressive types of nervous or emotional disorder." Another caution offered by the makers of these scales is that the results should be considered more significant for their relationship among the members of the local group than for their approach to standard norms. These limitations should be kept in mind in any consideration of the data presented. Even if we make all the allowance possible for the subjectivity of the schedules and for the unreliability of the Pearson formula, a careful study of the summary in Table VII must still convince us that the variations of these children in behavior is very great.

To continue, the correlations between the total behavior scores and the scores of each of the other traits studied are shown by the following *r*'s:

| | |
|---|---|
| Behavior—Freedom from physical defects | $r = .106 \pm .090$ |
| Behavior—Socio-economic status | $r = -0.144 \pm .088$ |
| Behavior—Mechanical ability | $r = -0.304 \pm .082$ |
| Behavior—Total music | $r = -0.079 \pm .093$ |
| Behavior—Total achievement | $r = -0.041 \pm .089$ |

It is patent that these correlations are all negative except for freedom from physical defects, but that none of the correlations is

[8] Published by the World Book Company, Yonkers-on-Hudson, New York, 1930.

TABLE VII

SCORES OF SIX PUPILS AND SUMMARY FOR FIFTY-EIGHT CHILDREN OF NORMAL INTELLIGENCE ON THE HAGGERTY-OLSON-WICKMAN BEHAVIOR RATING SCHEDULES

| Pupil | Schedule A | Schedule B Div. I | Schedule B Div. II | Schedule B Div. III | Schedule B Div. IV | Schedule B Total |
|---|---|---|---|---|---|---|
| AQ | 59 | 31 | 24 | 36 | 38 | 129 |
| AO | 42 | 25 | 15 | 40 | 46 | 126 |
| ... | ... | ... | ... | ... | ... | ... |
| AK | 52 | 17 | 21 | 20 | 15 | 73 |
| BB | 30 | 15 | 19 | 16 | 20 | 70 |
| ... | ... | ... | ... | ... | ... | ... |
| AA | 0 | 11 | 11 | 14 | 13 | 49 |
| BY | 0 | 12 | 8 | 12 | 13 | 45 |
| Mean | 20.94 | 18.7 | 15.1 | 22.7 | 21.0 | 75.64 |
| Median | 15.00 | 17.3 | 14.2 | 21.3 | 20.0 | 69.32 |
| Q. | 12.42 | 3.7 | 2.9 | 3.5 | 4.3 | 11.40 |
| S. D. | 19.96 | 5.3 | 3.9 | 6.0 | 7.5 | 18.89 |
| V. | 95.28 | 28.1 | 25.8 | 26.4 | 35.6 | 24.97 |
| Range | 78 | 23 | 16 | 27 | 35 | 84 |

high enough to have any considerable significance. We can safely say, then, that there is no significant relationship between behavior and any of these other traits.

Finally, the following pertinent conclusions may be drawn from a consideration of the data relative to behavior of these fifty-eight pupils of normal intelligence: (1) Pupils show a strong tendency toward individual differences, within themselves, even among the different phases of behavior; (2) pupils show a greater variation in emotional tendencies than in intellectual, physical, or social traits; and (3) there appears to be no significant relationship between the pupils' behavior scores and their scores in any of the other traits measured.

## INDIVIDUAL PROFILES

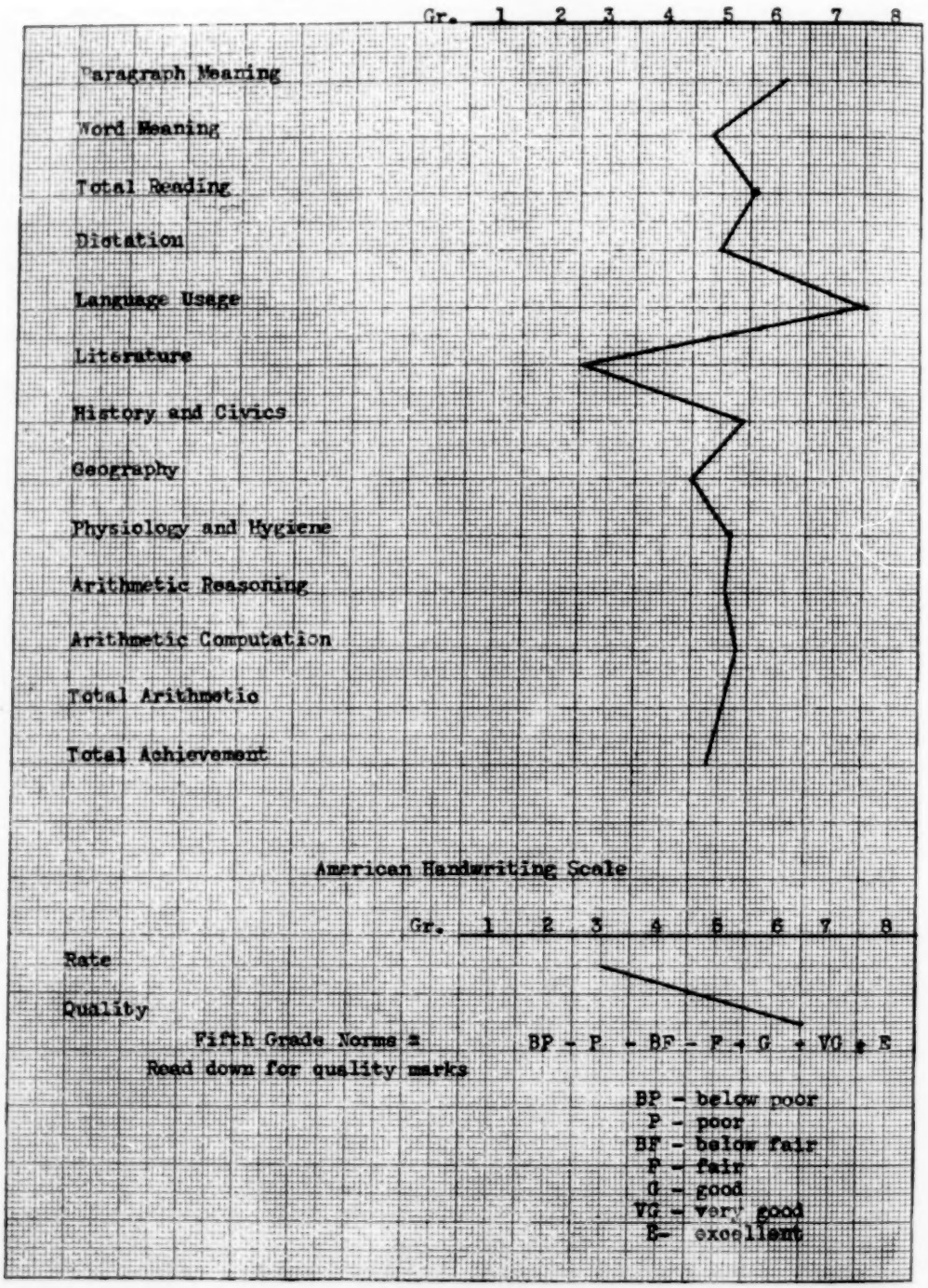
The data collected for all these pupils in the various traits studied were condensed into a series of profiles designed to give graphic pictures of each pupil's standing in the different characteristics. The arrangement of the data on the profiles was somewhat arbitrary, being so arranged as to group similar data on the same page and to use the same type of norms as nearly as possible. Each pupil's profile consists of two pages, on the first of which is the information relative to the New Stanford Achievement Tests and the American Handwriting Scale. These scores are compared with the standard grade norms. On the second page are shown the data for the Kwalwasser-Dykema Music Test, the Haggerty-Olson-Wickman Behavior Rating Schedules, the Stenquist Assembling Tests of Mechanical Ability, and the Sims Score Card for Socio-Economic Status; and these scores are compared with standard percentile or decile norms. All scores on the profiles are so arranged that high scores imply desirable traits. There is a separate profile of two pages for each of the fifty-eight pupils studied; but because of the lack of space only two of the individual profiles, showing the greatest and the least variation, are given here. Perhaps no comment on these profiles is necessary except to say that *of the fifty-eight pupils tested not one has all of his achievement scores within the scope of one school grade; not one even approximates an average score in all the music tests; and only one comes within twenty per cent of a uniform normal rating in behavior.*

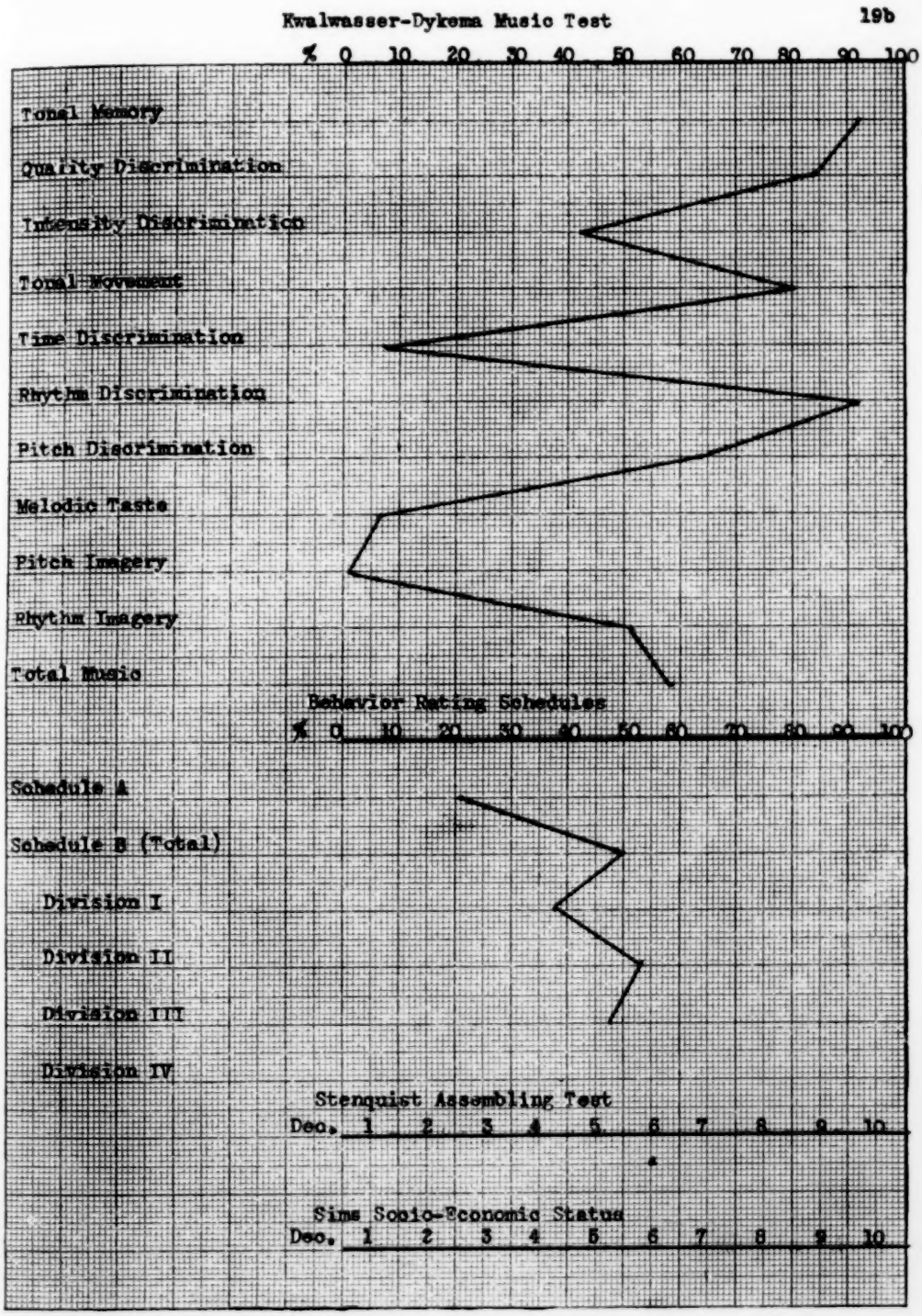
## RANKING PROFILES

As a final step in the endeavor to determine variations, the fifty-eight pupils were ranked, from 1 to 58, in height, weight, freedom from physical defects, socio-economic status, mechanical ability, musical ability, achievement, and behavior; and these rankings are assembled in Table VIII. It is to be noted that Pupil AA is eleventh from the top in height; i.e., there are ten taller than he; he is third from the top in weight, for two pupils weigh more than he; four have fewer physical defects than he; ten have a better socio-economic status than he; eighteen have better mechanical ability; fourteen are better in music; eight have a higher total achievement score; and one is rated by his teachers as having fewer undesirable behavior character-

Pupil AE

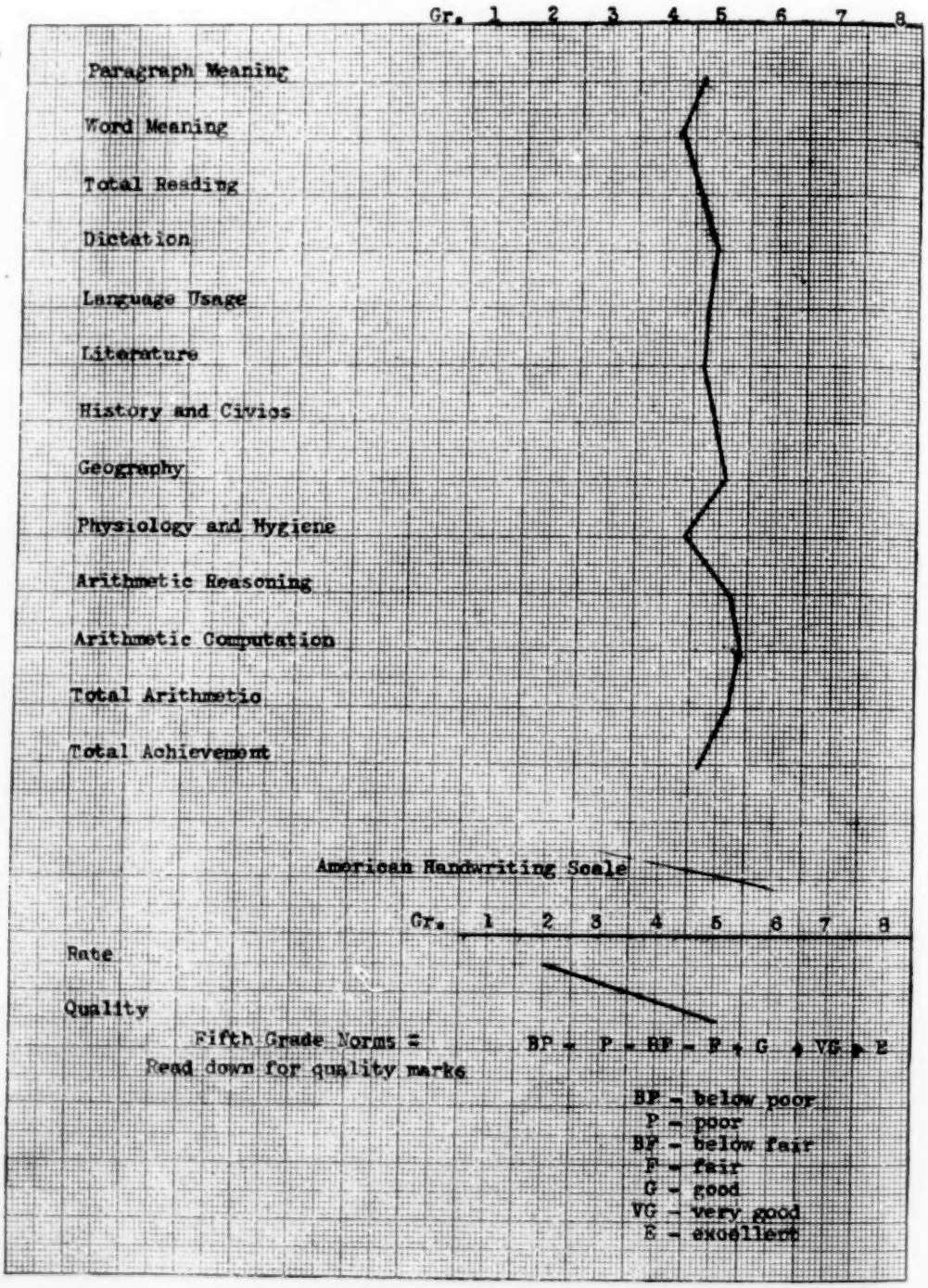
New Stanford Achievement Tests 19a




PUPIL AE
Kwalwasser-Dykema Music Test
19b
% 0 10 20 30 40 50 60 70 80 90 100
Tonal Memory
Quality Discrimination
Intensity Discrimination
Tonal Movement
Time Discrimination
Rhythm Discrimination
Pitch Discrimination
Melodic Taste
Pitch Imagery
Rhythm Imagery
Total Music
Behavior Rating Schedules
% 0 10 20 30 40 50 60 70 80 90 100
Schedule A
Schedule B (Total)
Division I
Division II
Division III
Division IV
Stenquist Assembling Test
Dec. 1 2 3 4 5 6 7 8 9 10
Sims Socio-Economic Status
Dec. 1 2 3 4 5 6 7 8 9 10

PUPIL BJ

New Stanford Achievement Tests 19c



PUPIL BJ 19d

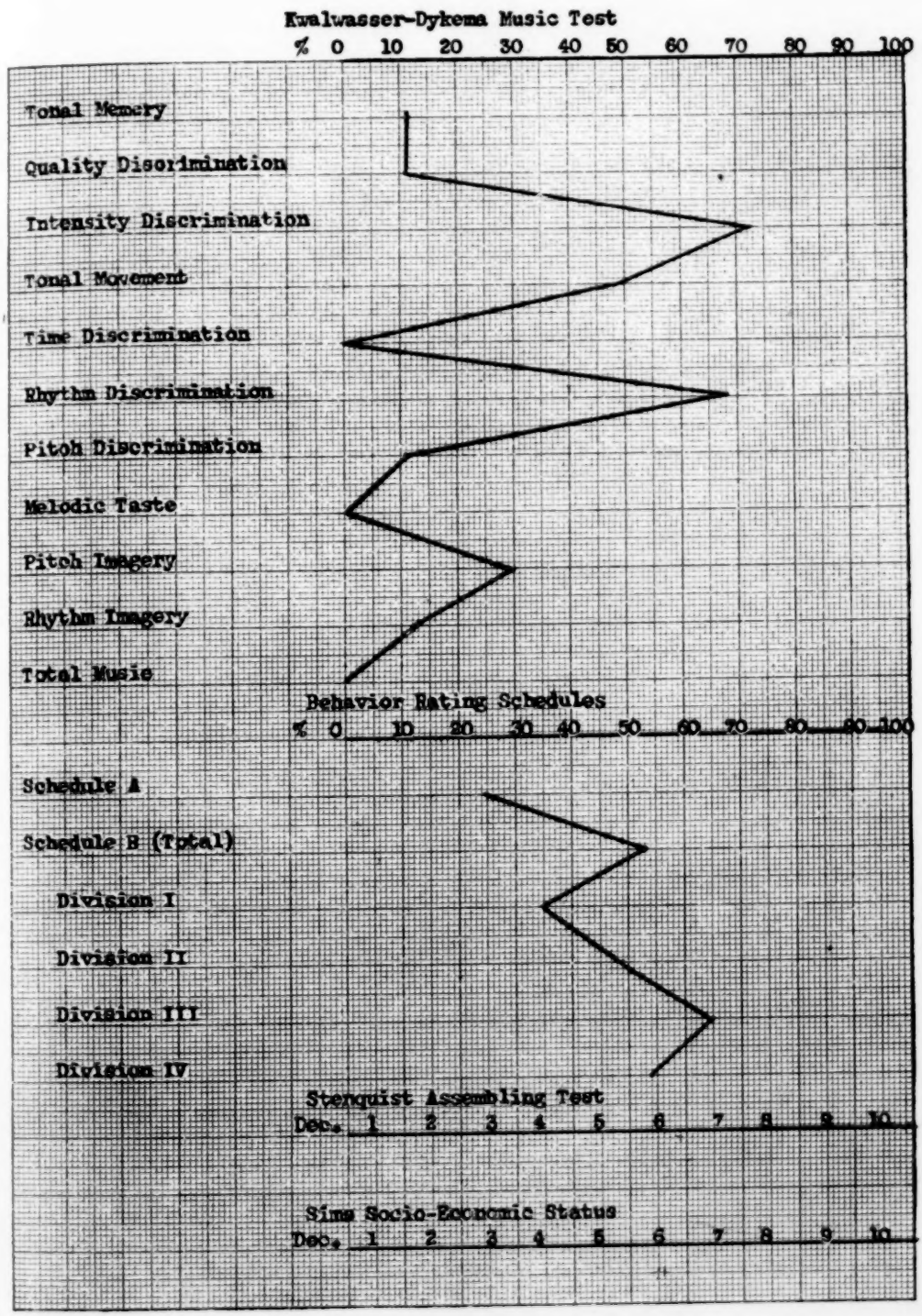

Kwalwasser-Dykema Music Test
% 0 10 20 30 40 50 60 70 80 90 100
Tonal Memory
Quality Discrimination
Intensity Discrimination
Tonal Movement
Time Discrimination
Rhythm Discrimination
Pitch Discrimination
Melodic Taste
Pitch Imagery
Rhythm Imagery
Total Music
Behavior Rating Schedules
% 0 10 20 30 40 50 60 70 80 90 100
Schedule A
Schedule B (Total)
Division I
Division II
Division III
Division IV
Stenquist Assembling Test
Dec. 1 2 3 4 5 6 7 8 9 10
Sims Socio-Economic Status
Dec. 1 2 3 4 5 6 7 8 9 10

TABLE VIII

RANKINGS OF FIFTY-EIGHT PUPILS IN THE EIGHT TRAITS STUDIED

| Pupil | Height | Weight | Freedom from Physical Defects | Socio-Economic Status | Mechanical Ability | Musical Ability | Achievement | Behavior |
|---|---|---|---|---|---|---|---|---|
| AA* | 11+ | 3 | 5 | 11 | 19 | 15 | 9 | 2 |
| AB | 7 | 9 | 1 | 21 | 34 | 31 | 17 | 48 |
| AC | 46 | 43 | 53 | 11 | 56 | 14 | 25 | 7 |
| AD | 46 | 50 | 30 | 27 | 52 | 31 | -- | 3 |
| AE* | 51 | 53 | 23 | 21 | 17 | 8 | 21 | 26 |
| AF* | 19 | 16 | 41 | 49 | 10 | 39 | 9 | 36 |
| AG | 40 | 39 | 10 | 29 | 50 | -- | 46 | 28 |
| AH | 19 | 28 | 10 | 34 | 15 | 42 | 1 | 14 |
| AI | 11+ | 36 | 52 | 45 | 7 | 3 | 35 | 35 |
| AJ | 19 | 39 | 30 | -- | -- | -- | 49 | 22 |
| AK | 7 | 28 | 7 | 17 | 2 | -- | 23 | 48 |
| AL | 7 | 18 | 10 | 34 | 22 | 35 | 41 | 12 |
| AM | 51 | 50 | 48 | 30 | 54 | 31 | 33 | 45 |
| AN | 19 | 35 | 51 | 30 | 55 | 35 | 42 | 19 |
| AO | 19 | 27 | 10 | 3 | 19 | -- | 36 | 55 |
| AP | 3 | 5 | 54 | 45 | 8 | 27 | 11 | 47 |
| AQ | 11+ | 6 | 48 | 55 | 12 | -- | 55 | 58 |
| AR | 28 | 37 | 23 | 10 | 3 | 6 | 2 | 54 |
| AS | 11+ | 21 | 45 | 30 | 6 | 10 | 32 | 16 |
| AT | 30 | 14 | 23 | 5 | 3 | 4 | 18 | 45 |
| AU | 11+ | 24 | 5 | 17 | 28 | 51 | 47 | 30 |
| AV | 30 | 27 | 30 | 13 | 36 | 6 | 22 | 33 |
| AW | -- | -- | -- | 13 | 17 | 44 | 50 | 10 |
| AX | 11+ | 16 | 10 | 2 | 39 | 6 | 30 | 51 |
| AZ | 7 | 11 | 30 | 49 | 52 | 42 | 54 | 44 |
| BA | 46 | 50 | 30 | 40 | 36 | 15 | 40 | 50 |
| BB | -- | -- | -- | 41 | 32 | -- | 28 | 37 |
| BC | 46 | 42 | 30 | 41 | 46 | 31 | 34 | 9 |
| BD | 30 | 33 | 39 | 6 | 28 | 40 | 53 | 40 |
| BE | 30 | 14 | 1 | -- | 12 | 12 | 52 | 24 |
| BF | 28 | 24 | 10 | 34 | 11 | 30 | 7 | 43 |
| BG | 3 | 11 | 10 | 6 | 23 | 28 | 4 | 27 |
| BH | 6 | 9 | 7 | 34 | 5 | 19 | 57 | 19 |
| BI | 11+ | 6 | 23 | 6 | 26 | 35 | 4 | 29 |
| BJ | 30 | 21 | 10 | 30 | 15 | 49 | 26 | 33 |
| BK | 51 | 49 | 30 | 20 | 39 | 19 | 56 | 6 |
| BL | -- | -- | -- | 13 | 1 | 25 | 16 | 37 |
| BM | 30 | 41 | 10 | 34 | 23 | 19 | 24 | 19 |
| BN | 30 | 21 | 45 | 52 | 32 | 46 | 51 | 42 |
| BO | 30 | 35 | 41 | 45 | 19 | 15 | 13 | 16 |
| BP | 18 | 18 | 30 | 17 | 48 | 23 | 39 | 12 |
| BQ | 17 | 37 | 23 | 54 | 45 | 12 | 20 | 39 |
| BR | 30 | 20 | 10 | 41 | 36 | 18 | 3 | 4 |
| BS | 30 | 33 | 30 | 45 | 46 | 45 | 44 | 10 |
| BT | 54 | 43 | 48 | 34 | 50 | 48 | 43 | 16 |
| BU | 41 | 24 | 7 | 13 | 43 | 19 | 28 | 23 |
| BV | 41 | 43 | 39 | 53 | 28 | 50 | 27 | 56 |
| BW | 46 | 47 | 10 | 20 | 35 | 1 | 6 | 5 |
| BX | 41 | 43 | 41 | -- | -- | -- | 12 | 57 |
| BY | 3 | 6 | 23 | 41 | 46 | 40 | 14 | 1 |
| BZ | 2 | 1 | 1 | 1 | 42 | 2 | 19 | 53 |
| CA | 55 | 53 | 1 | 17 | 40 | 46 | 37 | 24 |
| CB | 19 | 2 | 45 | 6 | 44 | 11 | 7 | 40 |
| CC | 40 | 47 | 21 | 27 | 26 | 28 | 45 | 7 |
| CD | 11+ | 13 | 10 | 49 | 31 | 25 | 31 | 31 |
| CE | 19 | 28 | 23 | 4 | 12 | 5 | 14 | 31 |
| CF* | 40 | 55 | 1 | 21 | 25 | 23 | 37 | 14 |
| CG | 1 | 3 | -- | 21 | 8 | 35 | 48 | 52 |
| Total Ranks | 55 | 55 | 54 | 55 | 56 | 51 | 57 | 58 |

*There is a ranking profile given in the present article for each pupil starred in this table.

+Where the same number appears two or more times in the same column the pupils are tied for the rank indicated.

--Where this symbol appears in the table it indicates that there is no record for this particular item.

istics. This table was arranged to rank the pupils, in all the traits, so that the lower the number in the ranking column the higher the rank in desirable characteristics; and contrawise the higher the number the more undesirable are the characteristics. There are fifty-eight pupils in our group, so the ranks are from 1 to 58. Thus a person ranking 1, 2, or 3 would be considered very good in the particular trait; one ranking 56, 57, or 58 would be very poor in the trait; and one ranking two or three points either up or down from 29 would be about average. In some instances two or more pupils are ranked the same in the same column; e.g., pupils AA, AQ, AU, AX, BI, and CD are all ranked 11 in height. This is because each of these pupils is 54 inches tall, and there are ten pupils who are taller. Wherever a dash appears in the table it indicates that there is no record for that pupil in that particular trait. Each dash in the table diminishes by one point the possible number of ranks in that column, and the total number of ranks in each column is shown at the bottom of the column. Consequently, fifty-five is the lowest possible rank in the height column, fifty-five the lowest in the weight column, fifty-four the lowest in the freedom from physical defects column, and so on to the last column (behavior) which is the only column that has the full fifty-eight possible ranks. A star following a pupil's letters in the table indicates that his rankings are also shown in an individual ranking profile. To conserve space only four ranking profiles are given here.

In this list, Pupil AA has the most nearly uniform rankings of any pupil. His rankings range from 2 in behavior to 19 in mechanical ability. AE is perhaps a good example of what may be expected of any pupil since he shows a wide variation in various characteristics. He ranks 8 in musical ability; somewhat above the mean in freedom from physical defects, socio-economic status, mechanical ability, achievement, and behavior; and rather close to the lower extreme in height and weight. AF has a range from 9 to 49, with four above the mean and four below. He reaches all the way from within nine points of the top in one trait to within nine points of the bottom in another. AX ranks below the median in three traits and above in five, being within one of the top in socio-economic status and within seven of the bottom in behavior. BM has a range of thirty-one, with three below average and five above. BO has a range of thirty-two, with four on each side of the median. BP, BQ and BR have similar (irregular) records. CB varies from 2 to 45, with five above mid-point and three below; CD has a high rank of 10 in freedom from physical defects, and a low one of 49 in socio-economic status, with four above the median and four below; and BJ ranks 10 in freedom from physical defects and 49 in musical ability, with four above median and four below.

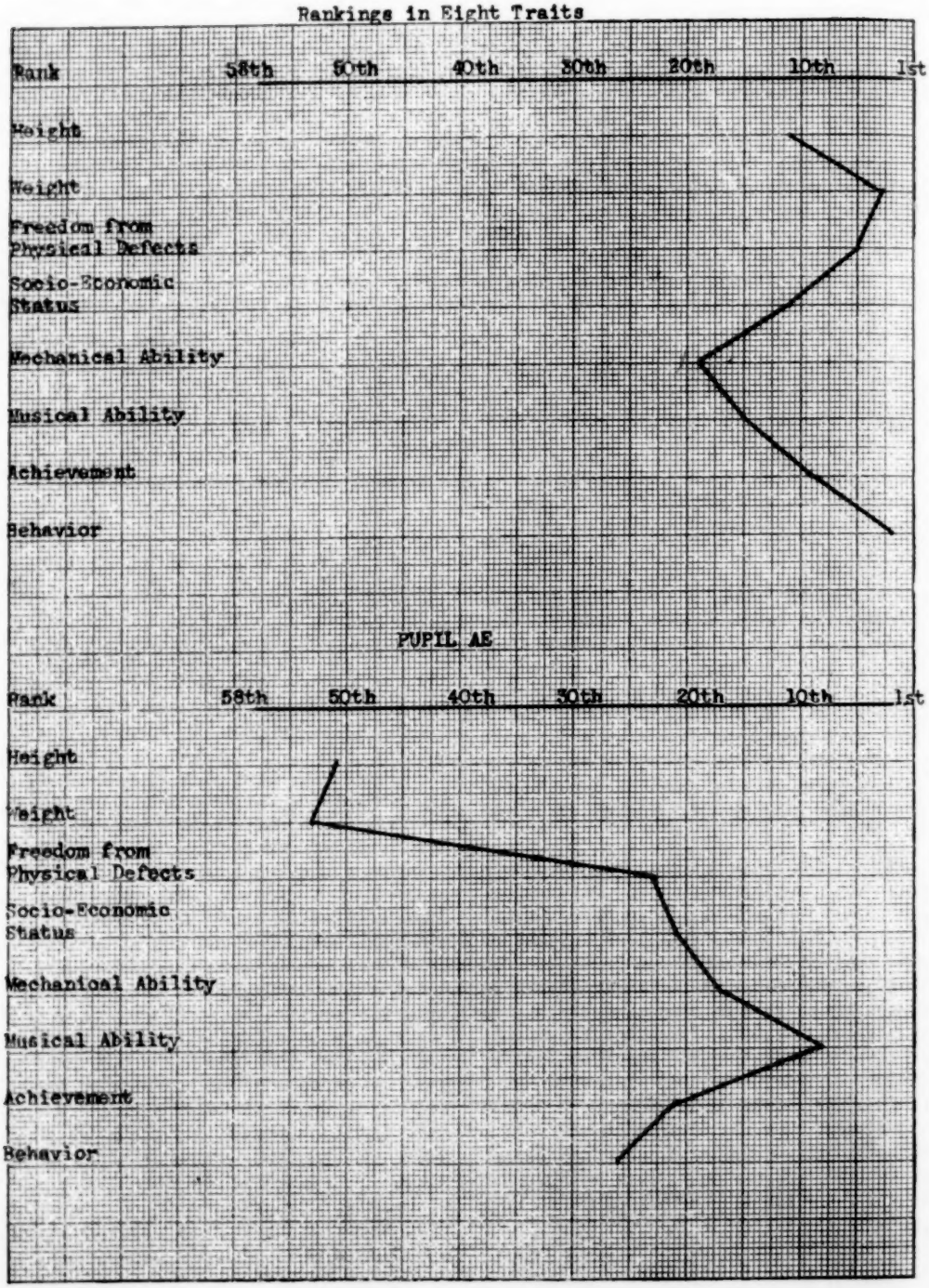
Even a casual glance at the table and the profiles is enough to convince one of the widespread differences of these pupils within themselves in their rankings. A more careful study will force one to the conclusion that there is little, if any, relationship among a pupil's rankings in these eight different traits —except in the case of height and weight.

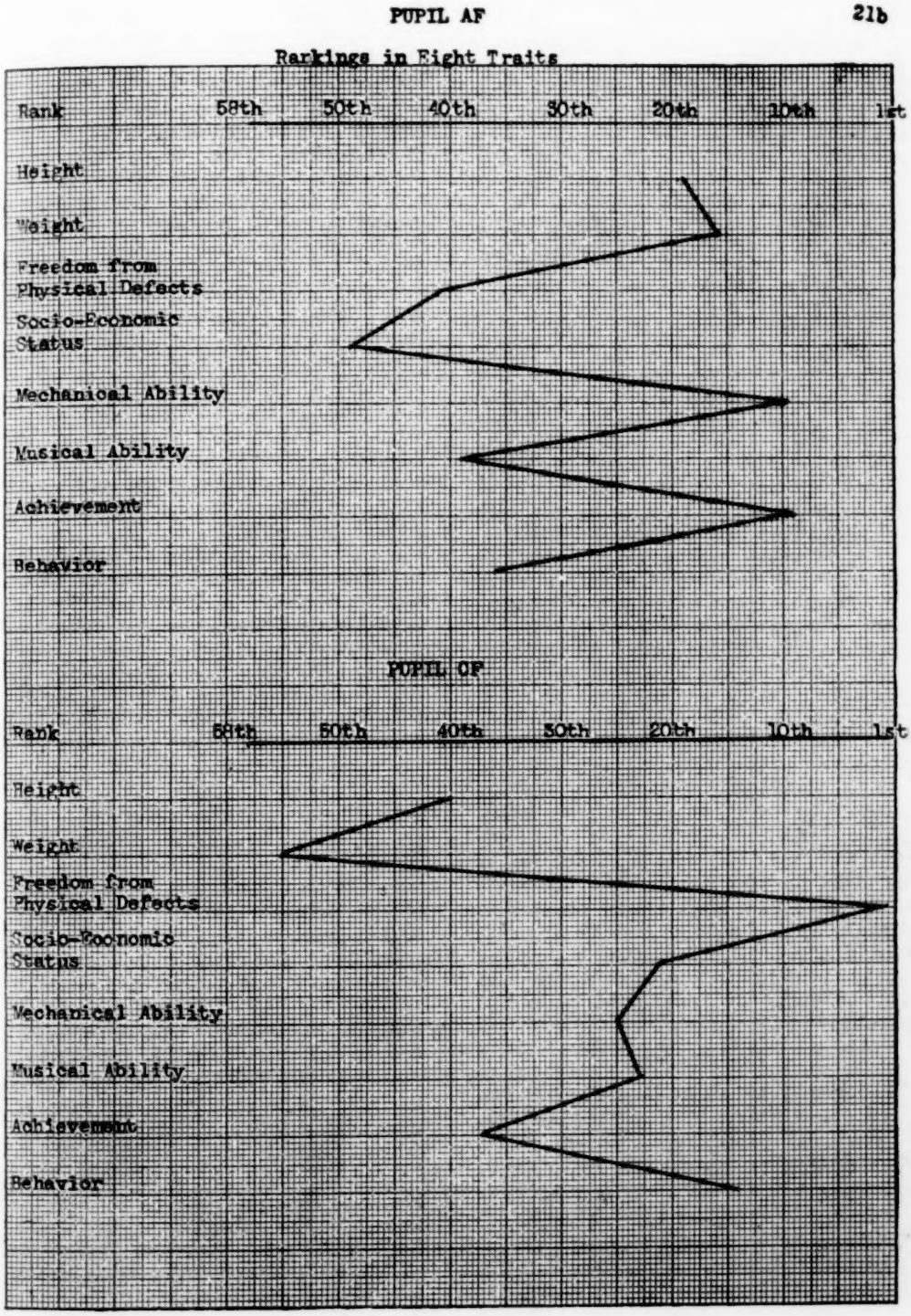
## Conclusions

Great individual differences among children of normal intelligence are shown to exist. Although children may be of normal intelligence as indicated by their I.Q.'s, it is clear that they are not as a rule homogeneous in any other traits that have been measured. In other words, it is impossible to predict where a pupil will stand in other traits from the fact that he is normal in intelligence. For that reason, if we wish to determine his capacities and accomplishments in the different elements of mechanics, music, the common branches, behavior, social standing, or economic status, we must study him directly in the trait in question. Any idea that children who are "normal" in intelligence are therefore likely to be uniform in any of the other traits is not borne out by the facts as found in this investigation. The data further show that no pupil who ranked high in any one trait was uniformly high in all other traits; that no pupil who was low in any one trait was uniformly low in all other traits; and that no pupil who was "average" in any one trait was uniformly average in all other traits. It is clear, therefore, that time, energy, and expense will be saved, and greater accuracy will be obtained by studying the child directly rather than by trying to predict a trait, or a group of traits, from his normal intelligence quotient.

PUPIL AA 21a

Rankings in Eight Traits




PUPIL AF
21b
Rankings in Eight Traits
Rank
58th
50th
40th
30th
20th
10th
1st
Height
Weight
Freedom from Physical Defects
Socio-Economic Status
Mechanical Ability
Musical Ability
Achievement
Behavior
PUPIL CF
Rank
58th
50th
40th
30th
20th
10th
1st
Height
Weight
Freedom from Physical Defects
Socio-Economic Status
Mechanical Ability
Musical Ability
Achievement
Behavior

## Implications

Teachers should be trained and given more time to work individually with their pupils. In addition, they should be made to appreciate the wide individual variations that are likely to occur in their children. They should be taught how to observe and diagnose the case of each individual child; how to determine what should be done after the diagnosis is made, and how to do the things that they have found necessary for his normal development. One difficulty in the past has been that teachers and administrators have believed in a superficial way that individual attention should be given to the children under their care, but they have not been trained sufficiently to give this individual attention. Again, many teachers are held for routine work, are overburdened with classes that are too large, or are limited to "standardized" outcomes set by school authorities which prevent them from giving the individual attention that is obviously needed.

The individual method used in the Winnetka Schools seems more sound in theory than do such schemes as the X, Y, Z grouping plan. The data gathered in this study furnished no evidence that there is an "average" child. The Winnetka plan takes care of the pupils individually in their regular school studies and also provides in the group activities for their social development. The X, Y, Z idea would place these fifty-eight pupils of "normal" intelligence in the "Y" group and assume that they are approximately homogeneous—which is not the case. Theoretically, then, some such plan as the one used at Winnetka provides for the individual needs of the child to a much greater extent than does the X, Y, Z method; since thereby children of "normal" intelligence, who differ among themselves as truly as do superior and dull children, are also given individual attention.

## Bibliography

"Bibliography on Mental and Physical Development," *Review of Educational Research*, III (April, 1933), 163–81.

"Bibliography on Mental and Physical Development," *Review of Educational Research*, VI (February, 1936), 124–52.

"Bibliography on Mental Hygiene and Adjustment," *Review of Educational Research*, VI (December, 1936), 545–54.

"Doctors' Theses Under Way in Education, 1933–1934," *Journal of Educational Research*, XXVII (January, 1934), 380–400.

"Doctors' Theses Under Way in Education, 1934–1935," *Journal of Educational Research*, XXVIII (January, 1935), 376–99.

"Doctors' Theses Under Way in Education, 1935–1936," *Journal of Educational Research*, XXIX (January, 1936), 384–406.

"Doctors' Theses Under Way in Education, 1936–1937," *Journal of Educational Research*, XXX (January, 1937), 370–96.

Freeman, Frank S. *Individual Differences: The Nature and Causes of Variations in Intelligence and Special Abilities.* New York: Henry Holt, 1934.

Leal, Mary A. "Personality Traits and Maturing of Children of Normal I.Q.," *Journal of Educational Research*, XXIII (March, 1931), 198–209.

Ledent, R. "Le Développement cérébral des enfants normaux," *Medical Scolaire*, XVI (1927), 3–7.

Marzi, A. "Sulle Attitudini Eidetiche visive nei Bambini Normali (Concerning Visual Eidetic Aptitudes in Normal Children)," *Scritti onore Kiesow*, 1933, pp. 169–177.

*Psychological Abstracts*, 1927–1937.

Rothbart, H. B. "Basal Metabolism in Children of Normal and of Sub-normal Intelligence: With Blood Cholesterol and Creatinine Values," *American Journal of Diseases of Children*, XLIX (March, 1935), 672–88.

Skinner, Charles E. *Educational Psychology*, pp. 387–582. New York: Prentice-Hall, 1936.

Strecker, Edward A. "Everyday Psychology of the Normal Child," *Mental Hygiene*, XVII (January, 1933), 65–81.

Terman, L. M., and Others. *Genetic Studies of Genius, Vol. I.* Stanford, California: Stanford University Press, 1925.

Washburne, Carleton W. "Burk's Individual System as Developed at Winnetka," *Twenty-Fourth Yearbook, National Society for the Study of Education, Part II.* Bloomington, Illinois: Public School Publishing Company, 1925.

Whiteside, S. "Spontaneity of Normal and Mentally Deficient Subjects in Selective Learning," *Proceedings, 1934*, pp. 344–84. Godfrey, Illinois: American Association on Mental Deficiency (Groves B. Smith, sec., Beverly Farm), 1934.

# SOLVING PROBLEMS OF ACTION: A PROPOSED PROCEDURE FOR THE SOLUTION OF PRACTICAL PROBLEMS

P. A. MAXWELL
*Peru State Teachers College, Nebraska*

Much of overt life activity consists of the expression of habit patterns that have come to be regarded as satisfactory modes of response to types of situations that recur frequently. However, adjustments often are made in such patterns of behavior as new elements appear in the stimulating situations. Sometimes these adjustments are made because an individual becomes disturbed over the efficacy of his acts and pauses to reflect, plan, and evaluate before acting overtly. At other times, especially in the lives of less thoughtful persons, such learnings may result from heedless trial and error overt responses or from uncritical acquiescence to suggestion which, like as not, is surcharged with emotional appeal.

Behavior patterns that are intelligent, wise, and in the long run most efficacious are those acquired by the reflective route. Intelligent, capable, self-directing persons are those who recognize the significant problems of action that confront them, who spare time when possible to study and reflect upon such problems, and who have become proficient in the techniques and mental processes involved in such study and reflection. To the degree, then, that intelligent, capable, self-directing persons are desired as products of education, guidance should be offered in the recognition and solution of significant types of practical or action problems. Obviously such guidance to be effective must provide the learners with adequate problem solving techniques. However, not a great deal has been done to perfect techniques for solving the action type of problem. To overcome this handicap and to provide his education classes with a procedure for solving teaching problems, the writer has developed the problem solving technique presented in this paper. For three years the assignments in all of his courses have been in the form of *"problacts,"* a term coined by contraction of the phrase "problem of action."

The term problact is significant, for the recognition of the distinctive nature of the action problem is an important step in the development of a satisfactory solution technique. In a problem of this type a person is disturbed and hesitates to act overtly, because he is uncertain what action is best or most appropriate. Thus the problact belongs to that class of problems which require the use of evaluation for their solution, and which may be distinguished from problems of the fact or science type. Abelson[1] and Monroe and Engelhart[2] have recognized the need for making this distinction in the field of educational research. The former refers to problems requiring a consideration of values as "evaluative problems," the latter as "problems of purposes." Dewey, too, has stressed such a distinction by devoting a chapter of *The Quest for Certainty* to a discussion of "The Construction of Good." The chapter contains the following statement:

"The time will come when it will be found passing strange that we of this age should take such pains to control by every means at command the formation of ideas of physical things, even those most remote from human concern, and yet are content with haphazard beliefs about the qualities of objects that regulate our deepest interests; that we are scrupulous as to methods of forming ideas of natural objects, and either dogmatic or else driven by immediate conditions in framing those about values."[3]

Another distinguishing feature of the problact is its comprehensive nature, which serves to distinguish it from other kinds of evaluative problems and from problems of the science type as well. Symonds[4] has recognized this in a classification of problems in which conduct problems are used to illustrate the "complete problem." Fact and simple value problems tend to arise incidentally in

[1] Harold H. Abelson, *The Art of Educational Research*, pp. 274–77. New York: World Book Co., 1933.
[2] Walter S. Monroe and Max D. Engelhart, *The Scientific Study of Educational Problems*, pp. 411–12. New York: Macmillan Co., 1936.
[3] John Dewey, *The Quest for Certainty*, p. 268. New York: Minton, Balch and Co., 1929.
[4] Percival M. Symonds, *Education and the Psychology of Thinking*, pp. 8–10. New York: McGraw-Hill, 1936.

the course of an attack upon a problact. Indeed, except in scientific research, concern for a question of truth or value is likely to be attributed to "idle curiosity," unless it arises in connection with a problact.

In general outline the proposed problact-solving procedure resembles that commonly employed in effective thinking about problems of all sorts. It involves: (1) an analysis and definition of the problem, (2) formulation of hypotheses, (3) testing or evaluation of the hypotheses, and (4) conclusion, or decision to follow the most promising hypothesis. It has been necessary, however, to devise many of the specific steps of the procedure in order to meet certain requirements peculiar to problact solving. This is true in particular of the evaluation technique, in which the hypotheses are compared in terms of their predicted consequences.[5] The principal difficulty with this technique has been encountered in the translation of the predicted consequences into *quantitative* expressions of value and the *synthesis* of the values thus expressed. Abelson[6] has held that "in most situations" this cannot be done, whereas Symonds[7] has suggested that when this point is reached one should turn his attention to something else or sleep on the matter, presumably to complete the evaluative process subconsciously. In developing the problact-solving procedure described below, it is believed that a measure of success has been attained in overcoming this difficulty.

Step 1. *Analysis of the Situation.* The first step in the proposed problact-solving procedure consists of a careful analysis of the problactic situation. This is accomplished by noting: (1) the *general field of activity* in which the situation is found—vocation, play, civic participation, home and family participation, the toilet, personal finance, etc.; (2) the particular *activity, function,* or *duty* involved—preparation of meals, curriculum adjustment, purchase of clothing, orchestra practice, and the like; (3) the *disturbing elements* of the situation—a recommendation from a respected source, a threat from a feared source, a sudden drop in personal efficiency, a stimulus arousing a new ambition, a stimulus that has come to serve as a danger signal, and other similar factors; and (4) *significant conditioning factors*—all elements of the situation likely to have an important bearing upon the nature of the outcomes when a response to the situation is made.

The analysis provides the groundwork for the other phases of the procedure. While in the main it should be worked out first, the search for significant conditioning factors may be continued after the formulation and evaluation of hypotheses have begun.

Step 2. *Formulation of Hypotheses.* Promising problact hypotheses in the form of plans of action may be created from elements of the past experience of the subject or of others. Often helpful suggestions may be obtained from appropriate literature. These may have the advantage of representing a number of differing points of view. Usually from two to four hypotheses are a satisfactory number, but to have sufficient *range* a set of hypotheses should represent all important points of view. To be *relevant* a hypothesis must promise to remove or ameliorate the disturbing elements of the situation.

Step 3. *Evaluative Comparison of the Hypotheses.* The discussion of this step is divided into three parts.

a. Formulation of Criteria of Value. When a satisfactory set of hypotheses has been formulated, the hypotheses are tested or evaluated in order to determine the relative merit or value of each. There are two possible methods of judging the merit of an act or series of acts. One consists of determining the degree of conformity of the act to principles or precepts that are considered sound. The other consists of determining the worth or desirability of the fruits or consequences of the act—*probable* consequences in the case of the hypothetical courses of action in problact solving. The latter method has been selected for our problact solving procedure. The precept method is considered to be but a makeshift procedure for use when time or other resources are insufficient to make the more thorough method of consequences practicable. The principal drawback to a general rule of action is its failure to make allowance for the many variations that are certain to exist in the class of situations to which the rule is supposed to apply.

The evaluation of a set of hypotheses consists, then, of comparing them in terms of the worth of their probable consequences. Now

[5] For an exposition of this method of evaluation see Symonds, *op. cit.*, Chapter XV on "Problem Solving" and Chapter XXIV on "Thinking About Conduct."

[6] *Op. cit.*, p. 282.

[7] *Op. cit.*, p. 159.

the worth or value of anything depends upon the amounts of desirable qualities it possesses. Therefore the *criteria of value* employed in comparing problact-solving hypotheses are variable qualities of objects likely to be affected by the hypothetical responses. This point is worthy of emphasis, for the common practice of referring to desired objects as values has been a serious obstacle to the development of effective evaluative techniques. Values always may be expressed as variable qualities of objects, that is, in measurable form and must be so expressed to serve as satisfactory problact-solving criteria.

Qualities employed as criteria of value in problact solving are *relevant* if their amounts depend somewhat upon the decision to be made. They are *significant* to the degree that they are prized for their own sake or condition other qualities so regarded. Good problact-solving criteria are both *specific* and *generic*. They are specific in that they are particular, definite values into which goodness in general may be analyzed. The more specific in this sense, the more accurate will be the predictions of hypothetical consequences. Good criteria are generic in the sense that each serves to measure the consequences of all the hypotheses formulated for a problact. Each is a quality of an object that will be present in some ensuing situation regardless of which hypothesis is followed. Like promising hypotheses, relevant and significant criteria may be discovered by recalling criteria used previously in the solution of similar problacts or by looking for suggestions in appropriate literature.

b. Organization of a Criterial Hierarchy. When a number of values have been selected to serve as criteria in the solution of a problact, it frequently is found that a means-end relationship exists between some of them. Indeed, series or strands each consisting of a number of the criteria usually are found to exist, the status of each member of a series depending somewhat upon the status of the preceding or *conditioning* member. For example, the solution of a problact concerned with the beautification of the grounds surrounding my home might involve the following strand:

| | | | | |
|---|---|---|---|---|
| **Height of beauty of the landscape in question** | — | **Friendliness of Neighbor X toward me** | — | **Strength of my personal prestige in the community** |

As a rule criterial strands to be most useful must begin with *immediate* criteria and extend through *intermediate* criteria to *ultimate* criteria. Immediate criteria are expressed in terms of the immediate consequences of the choice of a plan of action. They are qualities either of the hypotheses themselves or of objects affected directly by the hypotheses, and they should depend entirely or chiefly upon the choice made. The advantage of employing immediate criteria lies in the fact that the estimates of amounts likely to follow the supposed adoption of each hypothesis are relatively easy to make in their case; and, as will be seen later, such estimates provide a foundation for the prediction of consequences in terms of intermediate and ultimate criteria.

Intermediate criteria provide a medium through which the probable consequences of hypotheses may be traced to ultimate criteria. The latter are the final tests of value employed in the solution of a problact. Usually they are comprehensive values that are considered to be of paramount importance, and regarding the desirability of which there is a high degree of certainty. It is in this certainty that the value of the ultimate criteria lies, for by virtue of this feature the ultimate criteria provide a satisfactory basis for *weighting*, an essential step in the evaluation process.

When it is thought that all the significant strands of relevant criteria have been set up for the evaluation of a set of hypotheses, they may be combined into a unified evaluation instrument in the form of a *criterial hierarchy*. Usually it is found that some of the criteria employed are common to two or more of the strands. Such strands should be joined together by means of their common elements in order to bring into relief significant interrelationships.

c. Computation of Hypothetical Ratings and Value Indices. The actual comparison of problact-solving hypotheses begins with the estimation of the amount of each immediate criterion that is likely to exist following the supposed adoption of each hypothesis. The estimates are expressed as percentages of the optimal amount of each criterion. This is

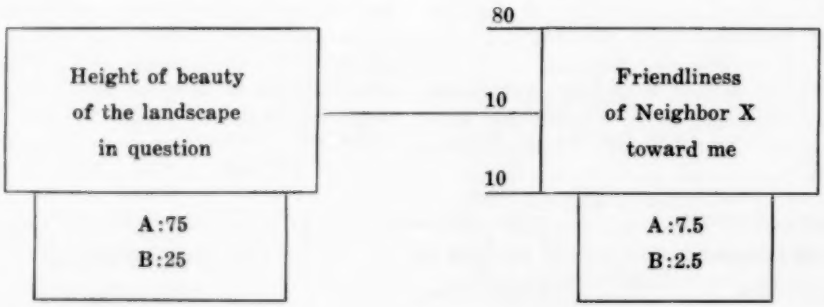
the ideal or best possible amount, the amount that is most favorable to the existence of desirable amounts of the dependent higher order criteria and that at the same time is possible under the limitations imposed by nature. The estimated percentages of the immediate criteria are spoken of as *immediate hypothetical ratings.*

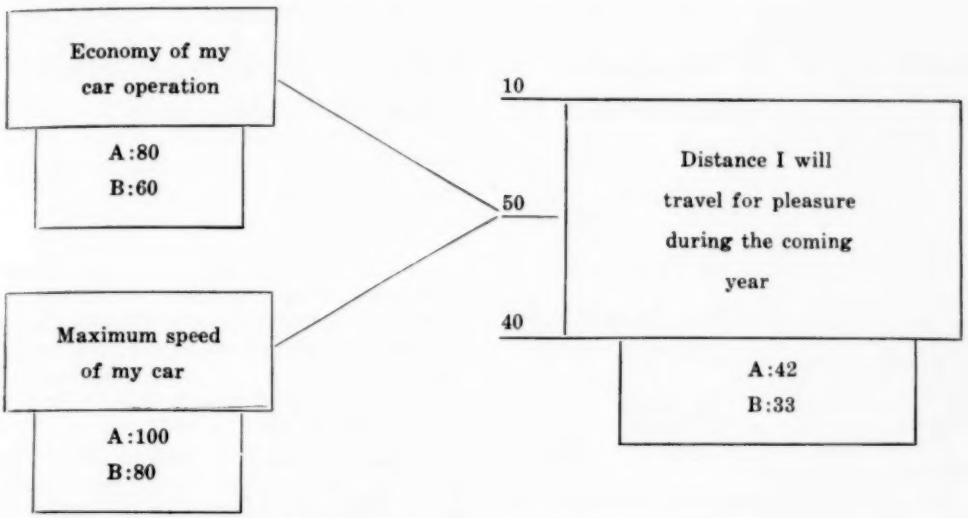
Hypothetical ratings in terms of intermediate and ultimate criteria are derived step by step from the immediate ratings, beginning with the intermediates, or at times ultimates, that are directly conditioned by the immediates. First an estimated optimal amount of each criterion is divided into three parts which are expressed in percentage of the optimum. The first part is a *predicted independent base,* the amount of the criterion that, owing to influences that are extraneous to the problact, will exist regardless of what is done by the problact solver in response to the problactic situation. The second part consists of one or more *maximal contingent increments,* which are the amounts that will be added to the independent base if the conditioning criteria are optimal in amount. The third part is the *minimal deficiency,* the portion of optimum that is unattainable due to influences that are extraneous to the problact.

The final step in computing the hypothetical ratings for each intermediate and ultimate criterion consists of determining the portion of the maximal contingent increment, or increments, that probably will accrue from each hypothesis. When an intermediate or ultimate criterion is considered to be in direct proportion to a conditioning criterion, the hypothetical increments may be computed by multiplying the maximal contingent increment by the respective hypothetical ratings of the conditioning criterion. This is the case in the following illustration, taken from the strand previously mentioned, in which the hypothetical ratings (A and B) of the conditioning criterion are 75 per cent and 25 per cent; the predicted independent base, 80 per cent; the maximal contingent increment, 10 per cent; the minimal deficiency, 10 per cent; and the computed hypothetical increments, or ratings, 7.5 per cent and 2.5 per cent:

In case there are two or more conditioners that vary proportionately with the dependent criterion and which influence it independently, partial hypothetical increments are derived from each one separately by the above procedure and then added together to form the complete hypothetical increments. However, if the influence exerted by one conditioning criterion depends upon the influence exerted by another, one maximal contingent increment and complete hypothetical increments are predicted for the two acting together, the hypothetical increments being estimates rather than computations. Below is an illustration of such a case, the predicted 42 per cent rating for Hypothesis A being based upon the combination of the two "A" ratings, 80 and 100, of the conditioning criteria, while the 33 per cent rating for Hypothesis B is based upon the corresponding combination of "B" ratings. The predictions in this case cannot be made separately, for the reason that speed may be so low that economy will not matter, and likewise economy may be so low that speed will not matter.

While experience has shown that criteria as a rule may be expressed in such form that dependents and conditioners vary in proportion, it is conceivable that other relationships might at times exist between them. In such cases the computation technique or method of estimate may be adapted to the relationship that does exist. Likewise it seems unnecessary as a rule to express criteria in nega-






tive form, but if that should be found desirable, the ratings may be expressed as negative quantities.

When hypothetical ratings have been derived for all the ultimate criteria of a criterial hierarchy, they are synthesized to form value indices for the several hypotheses. The first step in the synthesis consists of rating the ultimate criteria in importance by assigning a weight factor to each. The unit employed is the weight attached to the criterion ranking lowest in importance or value. By multiplying each weight factor by the hypothetical ratings of the corresponding ultimate criterion partial value indices are determined for the respective hypotheses. The sum of the partial indices determined for each hypothesis is the *index of value* for that hypothesis.

Step 4. *Conclusion.* Coming to a conclusion in problact solving consists merely of adopting the course of action—hypothesis—having the highest index of value. Obviously, the decision should be considered tentative, for the high ranking hypothesis may be considered good only in a relative sense. Whether it is the best possible or ideal policy has not been and cannot be determined.

Moreover, it should be clearly recognized that there are many chances for error in the solution of a problact. The confidence placed in a selected policy should depend upon the thoroughness with which it has been possible to analyze the situation and the certainty of the facts upon which the evaluations have been based. As a policy is carried into action its deficiencies likely will become apparent as they give rise to new problacts.

The procedure just outlined has been used for some time in education courses that are designed to prepare prospective teachers to solve teaching problacts. It is recommended for use in other fields of vocational education, guidance, social science, and in any other fields that are concerned with developing proficiency in the solution of action, conduct, or practical problems. Probably the procedure should be simplified somewhat if it is used at the elementary school level.

# METHODS OF TEACHING APPLIED TO JUNIOR BUSINESS TRAINING IN PHILADELPHIA*

HAROLD B. BUCKLEY
*Supervisor of Commercial Education,*
*Philadelphia Public Schools*

## I. INTRODUCTION

*The Problem.*—What are the relative differences in efficiency of four different methods of teaching when applied to junior business training in the junior high schools and evening high schools of Philadelphia?

The methods of teaching in this study are called:

1. The Blackboard and Discussion Method.
2. The Lesson-Sheet and Textbook-Reading Discussion Method.
3. The Blackboard and Textbook-Reading Discussion Method.
4. The Unit Assignment Method.

Three experiments were conducted to solve the problem of this study.

Experiment 1 compares Method 1 with Method 2 when both are applied to the teaching of individual lessons during a term.

Experiment 2 compares Method 1 with Method 3 when both are applied to the teaching of individual lessons during a term.

Experiment 3 compares Method 1 with Method 4 when both are applied to the teaching of the units, or series of related lessons, in the course of study during a term. In each experiment Method 1 is used with the control group.

The essential differences in the methods are as follows:

Method 1—The Blackboard and Discussion Method uses blackboards for developing in the classroom each of the four divisions of the individual lesson. The discussion method without textbook reading in the classroom is used for the presentation part of the lesson.

Method 2—The Lesson-Sheet and Textbook-Reading Discussion Method uses mimeographed lesson-sheets for developing in the classroom each of the four divisions of the individual lesson. The discussion method with textbook reading in the classroom preceding the discussion is used for the presentation part of the lesson.

Method 3—The Blackboard and Textbook-Reading Discussion Method uses blackboards as in Method 1. The discussion method with textbook reading in the classroom preceding the discussion is used as in Method 2.

Method 4—The Unit Assignment Method uses a unit or series of mimeographed lesson-sheets which are assigned to the pupils for independent work and textbook reading after the complete assignment has been discussed. This method differs from Method 2 in that the pupils work at their own rates of speed through the lesson-sheets. In this way the brighter pupils not only finish enrichment material at the end of each lesson-sheet, as they did in Method 2, but also are enabled to progress more rapidly than the other pupils through successive lessons. The bright pupils use the time saved to review the work of the unit as a whole.

The following is a detailed description of each method:

1. The Blackboard and Discussion Method. This method is used with the control group in each experiment. Each daily lesson in the classroom in which this method is used consists of four divisions:

A. Preparation or approach[1]

This, the first division of each individual lesson, the pupils are taught to call the *Self-reliance Work*. The device was originated by the writer in 1926, chiefly for the purpose of using to best advantage the first five minutes of a lesson period which are usually wasted. By this device, instead of waiting until the pupils are quietly seated in the classroom before giving orally the usual lesson-preparation questions, two questions are written on

* Doctorate dissertation, Temple University, 1937.

[1] John G. Kirk, Harold B. Buckley, and Mary A. Waesche, *Introduction to Business Manual and Key*, p. 11. Philadelphia: The John C. Winston Co., 1933. Pp. vi + 250.

the blackboard before the class arrives, as follows:

Self-reliance Work _____.

(1) A review question which requires for its answer a knowledge of the essentials of the previous lesson.

(2) A leading question which is apperceptive and which motivates the learning of the new lesson.

Each pupil independently answers question 1 in his notebook as soon as he takes his seat, while the teacher records attendance and does any other necessary work usually required at the beginning of a lesson period. The device is used in such a way that the class begins to finish the written work in about five minutes. The first pupil finishing question 1 automatically takes his work to the teacher and, if approved, that pupil becomes *Class Leader*, who at the front of the room does the following things:

a. gives place numbers to the next nine pupils who finish question 1 satisfactorily. The pupils record their place-number on a short line after the words Self-reliance Work _____.

b. conducts a discussion about the correct answer to question 1 and asks some pupil to write the answer on the blackboard for the class.

c. conducts a three-minute socialized recitation in which are used questions that the pupils have already prepared as part of their homework.

At this point, the teacher relieves the class leader and:

a. conducts the marking and collecting of homework by the pupil.

b. conducts the return of the pupils' homework of the preceding lesson.

c. discusses and demonstrates at the blackboard how to mark the writing of question 1 for one of fourteen common handwriting errors.

d. calls for answers to question 2 and uses the most desirable answers to lead to some business problem, the solution of which is the essential feature of the presentation part of the lesson.

B. Presentation or development[2]

This, the second division of each individual lesson, the pupils are taught to call the *Problem*, because it consists of solving some problem related to the pupils' business lives, such as "How should I write a business letter of inquiry?"

The Blackboard and Discussion Method is primarily a method of presentation in which the problem of the lesson is solved by a class discussion conducted by the teacher. In this method, textbook information relating to the solution of the lesson-problem is not read in the classroom. However, it should be remembered that in this method, as in all other methods in this study, textbook reading is included as part of the homework.

The discussion is followed by the development at the blackboard of a brief, informative outline, diagram, or model of a business form, which is copied by the pupils into their notebooks for future reference.

C. Application or conclusion in the classroom[3]

This, the third division of the lesson, the pupils are taught to call the *Practice Work*. It is an assignment which is related to the problem of the lesson and may consist of answering questions, preparing questions, filling in business forms, demonstrating a learned skill, or dramatizing a business situation. In order to recognize individual differences, the teacher encourages the "fast pupils" who finish ahead of the "average pupils" to assist the "slow pupils." When the "average pupils" finish, the teacher assigns one of the "fast pupils" to explain the finished work before the class. From this explanation, the "slow pupils" learn how to do the assignment correctly.

D. Application at home[4]

This, the fourth division, which is assigned in the classroom and done at home, the pupils are taught to call the *Homework*. It consists of three levels of work: (1) For a mark of F, fair, the pupil prepares one multiple-choice (four ending) objective question which deals with the problem of the lesson learned in the class-

[2] *Ibid.*, p. 12.
[3] *Ibid.*, pp. 13–14.
[4] *Ibid.*, pp. 14–15.

room and is to be used in the socialized recitation in the next lesson period; (2) for the mark of G, good, the pupil does the F assignment and reads the textbook information that is related to the problem of the lesson learned in the classroom, records his reading time, and prepares two more multiple-choice questions; and (3) for a mark of E, excellent, the pupil does the F and G assignments and some creative work showing complete mastery of the lesson. The homework is so arranged that it requires no more than fifteen minutes as a minimum to do assignments (1), (2), and (3) and to earn a mark of excellent.

The time allotment for a lesson period in the classroom is approximately as follows: (1) The first third of the period for the preparation or approach (i.e., Self-reliance Work); (2) the second third of the period for the presentation or development (i.e., Solving the Problem); and (3) the last third of the period for the application in the classroom (i.e., Practice Work) and assignment of the application at home (i.e., Homework). In this study the effective lesson period averages about thirty-nine minutes.

2. The Lesson-Sheet and Textbook-Reading Discussion Method. This method is the same as Method 1 except that:

A. a mimeographed sheet containing the four divisions of the complete daily lesson is given to each pupil at the beginning of each period. The lesson-sheet eliminates much blackboard writing by the teacher and notebook writing by the pupil, as in Method 1.

B. at the beginning of the presentation part of the lesson, textbook reading takes place before the discussion starts. It is silent reading and is arranged so that it requires for most lessons about two or three minutes and never more than five minutes of the time allotted for the presentation part of the lesson.

C. instead of copying an outline from the blackboard after the discussion, the pupils fill in a series of incomplete-type statements which are contained on the mimeographed sheet beneath the problem of the lesson. The words needed to fill in the incomplete statements really take the form of an outline.

3. The Blackboard and Textbook-Reading Discussion Method. This method is the same as Method 1 except that textbook reading takes place at the beginning of the presentation part of the lesson before the discussion starts, as in Method 2.

4. The Unit Assignment Method. The course of study during a term consists of a number of units or topics like "The Business Letter," "The Telephone," "Telegrams, Cablegrams, and Radiograms," etc.[5]

Each unit includes a series of individual lessons which are related to the same topic. For instance, the unit on The Business Letter includes five individual lessons, such as, "How should I write the parts of a business letter?" "How should I write a business letter of inquiry?" etc.

Each unit consists of five parts:

A. Pretest.—This consists of fifty multiple-choice (four ending) questions which deal with essential facts to be learned in the unit. A pretest in junior business training is possible because the facts relate to common daily business experiences.

B. Preview lesson.—This aims to give the pupils an orientation and overview of the unit to be studied and consists of the four divisions of an individual lesson.

C. Individual lessons.—These were previously described in detail.

D. Review lesson.—In this the presentation and application in the classroom consist chiefly of: (1) Arranging the finished work of the unit into a notebook report which includes a table of contents and cover sheet bearing the title of the unit and (2) conducting a socialized recitation on the unit.

E. Review test.—This might also be called the end test. It is the *same* test that was given as the pretest. In this study, the difference between the end-test score and the pretest score is spoken of as the Achievement or Number of Answers Learned.

The unit assignment method is similar to Method 2 except that all the lesson-sheets for the unit are distributed during the preview lesson. These lesson-sheets are discussed and then assigned as a whole unit for independent work in the classroom. The

[5] John G. Kirk, Harold B. Buckley, and Mary A. Waesche, *op. cit.*, pp. 22–35.

DESCRIPTIVE SUMMARY OF THE METHODS

| Activity | Method 1 | Method 2 | Method 3 | Method 4 |
|---|---|---|---|---|
| Introduction to the unit | —Pretest<br>—Preview lesson<br>(1) Blackboard outline<br>(2) Discussion without text reading in class | —Pretest<br>—Preview lesson<br>(1) Mimeographed lesson-sheet<br>(2) Discussion preceded by text reading in class | —Pretest<br>—Preview lesson<br>(1) like Method 1<br>(2) like Method 2 | —Pretest<br>—Preview lesson<br>(1) Unit set of mimeographed lesson-sheets<br>(2) like Method 1 |
| Presentation of new material | (1) Preparation questions and Socialized Recitation<br>(2) Problem stated on blackboard<br>(3) Discussion without text reading in class<br>(4) Blackboard outline copied into notebooks | (1) like Method 1<br>(2) Problem stated on lesson-sheet<br>(3) Discussion preceded by text reading in class<br>(4) Filling in incomplete statements on lesson-sheets | (1) like Method 1<br>(2) like Method 1<br>(3) like Method 2<br>(4) like Method 1 | (1) Preparation questions only<br>(2) like Method 2<br>(3) No discussion, but text reading independently<br>(4) like Method 2 except it is done independently |
| Fixation of new material<br>A. In class | (1) Answering questions, oral or written, prepared by teacher or assigned in textbook<br>(2) Preparing objective questions<br>(3) Filling in business forms<br>(4) Demonstration of specified skills<br>(5) Dramatization of business situations | (1), (2), (3), (4), and (5) like Method 1 | (1), (2), (3), (4), and (5) like Method 1 | (1), (2), and (3) like Method 1; (4) and (5) omitted |
| B. At home | (1) Preparing objective questions<br>(2) Reading textbook<br>(3) Collecting and preparing illustrative material | (1), (2), and (3) like Method 1 | (1), (2), and (3) like Method 1 | (1), (2), and (3) like Method 1 |
| Conclusion of the unit | —Review lesson<br>(1) Preparing notebook report<br>(2) Socialized recitation<br>—End test | —Review lesson like Method 1<br>—End test | —Review lesson like Method 1<br>—End test | —Review lesson like Method 1<br>—End test |

number of days allotted for independent work on the unit is equal to the number of individual lessons in the unit.

## TIMELINESS

The purpose of the study is to solve certain administrative and supervisory problems encountered by the writer as supervisor of commercial education in the day and evening high schools of Philadelphia.

The purposes of Experiment 1 are:

(1) To compare our present method of teaching individual lessons in the day junior high schools by the blackboard and discussion method with our present method of teaching them in the evening high schools by the lesson-sheet and textbook-reading discussion method.

(2) To determine the advantages, if any, of lesson-sheets which have been used in junior business training in the evening high schools of Philadelphia for the past five or more years.

The purposes of Experiment 2 are:

(1) To compare our present method of teaching individual lessons in the day junior high schools by the blackboard and discussion method, which omits

textbook-reading in the classroom, with the proposed method of teaching them by the blackboard and textbook-reading discussion method, which includes textbook-reading in the classroom before the discussion.

(2) To determine whether it is educationally better to include or exclude the reading of the textbook in the classroom.

The purpose of Experiment 3 is:

To compare our present method of teaching series of related individual lessons or units by the blackboard and discussion method, led by the teacher, with teaching them by the so-called unit assignment method.

### Previous Research

"A statement by a man prominent in the field of commercial education summarizes very well, indeed, the status of research in this field. The statement was made in 1926 by Mr. E. W. Barnhart at a commercial education research conference held at the University of Iowa. He said at that time: 'What are the solved and the unsolved problems in commercial education today? Frankly, none appears to be solved.' It is fair to say that this statement is as true today as when first it was made."[6]

The reason may be found in the following statement: "Studies of the professional preparation of business teachers in five states show that from twenty-nine to sixty-three per cent of teachers of business subjects hold baccalaureate degrees as compared with seventy-eight per cent of academic teachers who hold degrees.

"This comparatively low showing on the part of the business teachers has not been entirely their fault, but has been due in part to limited opportunity for professional training and improvement."[7]

Most of the research studies in commercial education have been: (1) Occupational surveys; (2) job analyses; and (3) school surveys. Few studies have dealt with methods of teaching commercial subjects and practically all of these have dealt with methods of teaching shorthand, typewriting, and bookkeeping. No study has been found which deals with methods of teaching junior business training.

[6] W. R. Odell, "Research in Commercial Education," *Journal of Business Education*, X (April, 1935), 11.

[7] Benjamin R. Haynes and Jessie Graham, *Research in Business Education*. Los Angeles: University of Southern California: Claude C. Crawford, 1932. Pp. 232.

## II

### Experiment No. 1

#### *The Problem*

What are the relative differences in efficiency of Method 1 (the blackboard and discussion method), compared with Method 2 (the lesson-sheet and textbook-reading discussion method), when the methods are applied to teaching individual lessons in junior business training?

#### *Procedure*

To solve the problem, the parallel-group or equivalent-group method of experimentation is used and Method 1 is used with the control group.

### Collection of Data

#### *Selection of Teachers and Schools*

It was desirable but not possible to conduct the experiment in evening high schools because each teacher has only one class. To overcome this disadvantage, it was decided that teachers of junior business training in both day junior high and evening high schools should be selected. Four of the five selected teachers had taught in both types of schools and had used both of the methods included in the experiment.

The experiment was conducted in four junior high schools where it was possible to arrange with the principal and roster chairman to assign each selected teacher two classes of the upper eighth grade. The schools selected were in various sections of Philadelphia in order that pupils with possibly individual and community differences might be included in the experiment.

#### *Time and Duration of Experiment*

Both methods of teaching were applied daily by each of the five selected teachers in two different eighth grade classes during the Spring term of eighteen weeks from February to June, 1935.

#### *Directions to Teachers*

Each selected teacher was given: (1) A typewritten sheet of directions which dealt with the methods of teaching each individual lesson, as shown in the appendix, (2) a

mimeographed sheet for each class for reporting I.Q.'s, and results of seven junior business training tests (one term and six unit tests) given as pretests before teaching and given again as review or end tests, etc., as shown in the appendix, and (3) specially prepared three-by-five cards for recording I.Q., results of each test, and other information about each pupil, as shown in the appendix.

*Data Provided by the Teachers*

Each teacher submitted to the writer: (1) A mimeographed sheet for the control class and one for the experimental class concerning I.Q.'s, results of tests, etc., and (2) a specially prepared three-by-five record card for each pupil in the experiment. A total of 420 pupil-record cards were received from the five teachers conducting the experiment, as shown in Table I.

In Table II are reported the means and standard deviation for reading scores, I.Q.'s, and pretests for the control group on which was used Method 1 and for the experimental group on which was used Method 2. From these, it may be seen that the differences of the means in each case are less than the probable errors of the differences.

*Summarizing the Results of the Junior Business Training Tests*

Since each junior business training test dealt with knowledges, skills, and attitudes that were separate, distinct, unrelated, and independent of those in the other junior business training tests, it was possible to obtain a total test score for each pupil based on 356 items. Each of these items was answered twice, once in a pretest and once again when the same test was given as a review or end test. However, since some of the pupils were absent for a given test and matched with another pupil also absent in the same test, there were averages of 323 pretest items and 323 end test items answered by each pupil. In other words, the average number of questions answered was 646.

TABLE I

SOURCES OF DATA FOR EXPERIMENT NO. 1

| School No. | Grade | No. Teachers | No. Classes | Pupils, Method 1 No. | Pupils, Method 1 Mean I.Q. | Pupils, Method 2 No. | Pupils, Method 2 Mean I.Q. |
|---|---|---|---|---|---|---|---|
| 1 | 8B | 1 | 2 | 44 | 110 | 36 | 111 |
| 2 | 8B | 2 | 4 | 90 | 104 | 79 | 111 |
| 3 | 8B | 1 | 2 | 41 | 107 | 42 | 109 |
| 4 | 8B | 1 | 2 | 42 | 115 | 46 | 104 |
| | | 5 | 10 | 217 | 108 | 203 | 109 |

## TREATMENT OF DATA

*Equating the Groups*

After the 420 pupil-records were checked for accuracy with the information contained on the class-report sheets, certain pupil-records were selected in order further to equate the groups which had already been equated as to grade, mean chronological age, and teacher. This further equating of the groups resulted in 320 pupil-records or seventy-six per cent of the total group in the experiment. Each pupil in the experimental group was matched with another pupil in the control group who had the:

(1) Same number of junior business training tests, both pretests and review or end tests.
(2) Same standard score on the Philadelphia Reading Test within 1 point.
(3) Same I.Q. within five points. (The Philadelphia Mental Ability Test was used.)

Distributions were prepared to show scores in pretests. Similar and comparable distributions show scores for end tests. For each pupil the difference between the sum of his end test scores and the sum of his pretest scores was obtained. This difference was called the *number of answers learned.*

Since the number of answers learned was dependent upon the number of questions not answered in the pretests, it was necessary to find what per cent of the wrong answers in the pretests were answered correctly in the end tests. This was found by the following formula:

$$\text{Per Cent of Possible Gain} = \frac{\text{Number of Answers Learned}}{\text{Total Number of Questions in Pretests Minus Total Score in Pretests.}}$$

In order to discover what effect, if any, the different methods might have had for pupils with high and low reading scores, these data were studied for high-, low-, and middle-reading score groups as well as for the total group.

## Findings

Tables II to V summarize the results of Experiment No. 1.

Table II shows that for the total group there is no significant difference in the means of the control and experimental groups either in the number of answers learned or in the per cent of possible gain. The control group or Method 1 showed a mean number of answers learned of 66.81, whereas the experimental group or Method 2 showed 67.81. This difference of 1.00 additional answer learned in favor of the experimental group is not significant in view of the fact that the probable error of the difference is 1.72 answers. The difference in mean per cent of possible gain is 1.66 per cent points in favor of the experimental group. Again, this is not a significant difference, since the probable error of the difference is 0.65 per cent points.

However, it should be noted that in both of these measures of the success of pupils, the differences, while small, are in favor of the experimental groups.

Table III, which deals with the pupils in the middle fifty-two per cent of the reading score distribution, confirms the findings for the total group. It gives additional evidence of no significant differences in the means of the control and experimental groups in the number of answers learned or in the per cent of possible gain.

It should be noted that in both of these measures of success of pupils the difference of means of the number of answers learned is 0.71 in favor of the control group, whereas the difference of means of the per cent of

Table II

Comparison of Results for All Pupils

| | Method 1 Control Group (N=160) | | | Method 2 Exp. Group (N=160) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Phila. Reading Test Standard Score | 12.0 | 3.17 | 1.69 | 12.0 | 3.20 | 1.70 | 0.0 | 2.41 |
| Phila. Mental Ability Test I. Q. | 108.9 | 11.1 | 0.59 | 108.9 | 11.1 | 0.59 | 0.0 | 0.84 |
| Pretests | 167.38 | 33.10 | 1.77 | 168.75 | 35.10 | 1.87 | (X) 1.37 | 2.57 |
| End Tests | 234.19 | 40.70 | 2.17 | 236.56 | 48.50 | 2.58 | (X) 2.37 | 3.37 |
| Number of Answers Learned | 66.81 | 22.40 | 1.19 | 67.81 | 23.40 | 1.25 | (X) 1.00 | 1.72 |
| Per cent of Possible Gain | 43.40 | 12.10 | .64 | 45.06 | 12.20 | .65 | (X) 1.66 | .65 |

Table III

Comparison of Results of Pupils in the Middle 52 Per Cent of the Reading Score Distribution

| | Method 1 Control Group (N= 84) | | | Method 2 Exp. Group (N= 84) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Pretests | 163.69 | 28.60 | 2.10 | 170.95 | 34.40 | 2.53 | (X) 7.26 | 3.29 |
| End tests | 231.90 | 37.60 | 2.77 | 238.45 | 48.30 | 3.55 | (X) 6.55 | 4.50 |
| Number of Answers Learned | 68.21 | 22.50 | 1.65 | 67.50 | 22.00 | 1.62 | (C) .71 | 2.30 |
| Per cent of Possible Gain | 43.22 | 11.40 | .84 | 44.17 | 13.30 | .98 | (X) .95 | 1.29 |

possible gain is 0.95 per cent points in favor of the experimental group.

Table IV shows that, for the pupils in the upper twenty-six per cent of the reading score distribution, there is no significant difference in the means of the control and experimental groups either in the number of answers learned or in the per cent of possible gain.

In both of the measures of the success of pupils, the difference in mean number of answers learned (7.03) favors the experimental group and is approximately twice the probable error of the difference of the means (3.54). The difference in mean per cent of possible gain (1.70 per cent points) also favors the experimental group and is greater than the probable error of the difference of the means (1.34).

Table V shows that for the pupils in the lower twenty-two per cent of the reading score distribution there is no significant difference either in the mean number of answers learned or in the mean per cent of possible gain. However, it should be noted that the difference in both measures favors the control group.

### Summary of Significant Findings

Figure 1 shows the percentile distributions of answers learned expressed as per cents of possible gain. It confirms the findings of the preceding tables. For instance, Figure 1 shows slightly higher results in Method 2 for pupils in the upper part of the distribution of percentage of possible gain. This confirms the findings of Table IV, which shows that the pupils with the higher reading scores had higher results with Method 2. While the evidence is not conclusive on this point, there is some reason for thinking that pupils at the upper end of the reading score distribution and those at the upper end of the distribution of percentage of possible gain are more successful with Method 2. To some degree these are the same pupils.

Figure I shows slightly higher results for the lower part of the distribution of percentages of possible gain in Method 1. This confirms the data of Table V, which shows that pupils in the lower end of the reading score distribution are more successful with Method 1.

**Table IV**

**Comparison of Results of Pupils in the Upper 26 Per Cent of the Reading Score Distribution**

| | Method 1 Control Group (N= 41) | | | Method 2 Exp. Group (N= 41) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Pretests | 189.10 | 37.30 | 3.93 | 180.85 | 31.30 | 3.30 | (C) 8.25 | 5.11 |
| End Tests | 252.56 | 46.90 | 4.94 | 251.34 | 47.80 | 5.04 | (C) 1.22 | 7.06 |
| Number of Answers Learned | 63.46 | 23.50 | 2.48 | 70.49 | 23.90 | 2.52 | (X) 7.03 | 3.54 |
| Per Cent of Possible Gain | 47.93 | 12.70 | 1.34 | 49.63 | 12.70 | 1.34 | (X) 1.70 | 1.34 |

**Table V**

**Comparison of Results of Pupils in the Lower 22 Per Cent of the Reading Score Distribution**

| | Method 1 Control Group (N= 35) | | | Method 2 Exp. Group (N= 35) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Pretests | 150.71 | 23.50 | 2.68 | 149.86 | 33.20 | 3.78 | (C) .85 | 4.63 |
| End Tests | 218.14 | 30.90 | 3.52 | 214.71 | 41.70 | 4.76 | (C) 3.43 | 5.92 |
| Number of Answers Learned | 67.43 | 20.40 | 3.32 | 64.85 | 22.20 | 2.53 | (C) 2.58 | 4.17 |
| Per Cent of Possible Gain | 38.71 | 11.20 | 1.28 | 38.43 | 11.70 | 1.33 | (C) .28 | 1.84 |

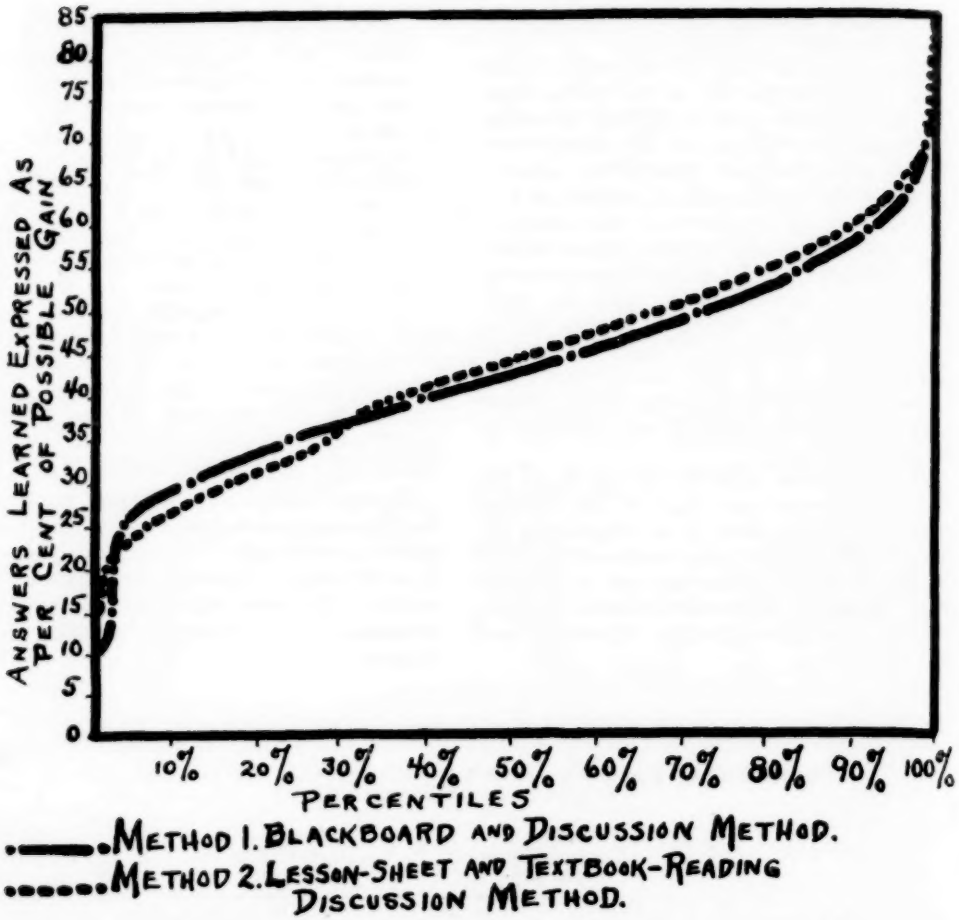



FIG. 1.- EXPERIMENT NO. 1, PERCENTILE DISTRIBUTIONS OF ANSWERS LEARNED EXPRESSED AS PER CENTS OF POSSIBLE GAIN

## III

### EXPERIMENT NO. 2

*The Problem*

What are the relative differences in efficiency of Method 1 (the blackboard and discussion method), compared with Method 3 (the lesson-sheet and textbook-reading discussion method) when the methods are applied to teaching individual lessons in junior business training?

*Procedure*

The procedure is the same as in Experiment No. 1, except where differences are noted.

### COLLECTION OF DATA

*Selection of Teachers and Schools*

Three of the five selected teachers had taught junior business training in both day and evening high schools and all of the teachers had experience with both of the methods of teaching in this experiment.

*Directions to Teachers*

The typewritten sheet of directions which dealt with the methods of teaching each individual lesson was similar to that shown in the appendix, except that the procedure for Method 3 was used instead of Method 2. Method 3 differed from Method 1 only in

that the class discussion in the problem-solving division was preceded by textbook reading in the classroom.

*Data Provided by the Teachers*

A total of 396 pupil-record cards were received from the five teachers, as shown in Table VI.

## Treatment of Data

*Equating the Groups*

Further equating of the groups resulted in 250 pupil-records or sixty-three per cent of the total group in the experiment. Each pupil in the experimental group was matched with another pupil in the control group who had the: (1) same number of junior business training tests, both pretests and end tests; and (2) same I.Q. within two points.[8]

[*] This teacher was assigned three 8B classes. In two of the classes he conducted Experiment No. 1 and in the third class he conducted Experiment No. 2. Otherwise, Experiment No. 2 was conducted in schools different from those in Experiment No. 1.

[8] In Experiments 2 and 3, the I. Q. was used because higher correlation was found to exist between I. Q. and per cent of possible gain than between reading score and per cent of possible gain. The correlation coefficients are as follows for the control and experimental groups in Experiment No. 1:

| | *Control* | *Experimental* |
|---|---|---|
| Between reading score and per cent of possible gain | .37 ± .05 | .36 ± .05 |
| Between I.Q. and per cent of possible gain | .50 ± .04 | .57 ± .04 |

From Table VII it may be seen that each group was equal as to mean I.Q. and that the difference of the mean of the pretests, although it was 3.64 answers in favor of the experimental group, was not significant. The difference was only slightly larger than the probable error of the difference (3.03).

*Summary of the Results of the Junior Business Training Tests*

The average number of questions answered by each of the 250 pupils was 666 questions, of which 333 were on pretests and 333 on end tests.

## Findings

Tables VII to X summarize the results of Experiment No. 2.

Table VII shows that for the total group there is no significant difference in the means of the control and experimental groups either in the number of answers learned or in the per cent of possible gain. The mean number of answers learned by the control group or Method 1 was 79.84 and by the experimental group or Method 3 was 74.28. This difference of 5.56 additional answers learned favors the control group, but is not significant in view of the fact that the probable error of

### Table VI

Sources of Data for Experiment No. 2

| School No. | Grade | No. Teachers | No. Classes | Pupils, Method 1 No. | Pupils, Method 1 Mean I.Q. | Pupils, Method 3 No. | Pupils, Method 3 Mean I.Q. |
|---|---|---|---|---|---|---|---|
| 1 | 8B | 1 | 2 | 49 | 100 | 51 | 100 |
| | | | 1 | 23 | 100 | -- | --- |
| 2 | 8B | 2 | 2 | 40 | 108 | 32 | 104 |
| | | | 2 | 35 | 93 | 40 | 114 |
| *3 | 8B | 1 | 1 | -- | --- | 40 | 102 |
| 4 | 8B | 1 | 2 | 45 | 108 | 41 | 108 |
| | | 5 | 10 | 192 | 102 | 204 | 105 |

### Table VII

Comparison of Results for All Pupils

| | Method 1 Control Group (N=125) Mean | S. D. | P. E. Means | Method 3 Exp. Group (N=125) Mean | S. D. | P. E. Means | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| Phila. Mental Ability Test I. Q. | 104.3 | 10.65 | 0.64 | 104.3 | 10.65 | 0.64 | 0.0 | 0.95 |
| Pretests | 165.80 | 36.60 | 2.21 | 169.44 | 34.50 | 2.08 | (X) 3.64 | 3.03 |
| End tests | 245.64 | 49.50 | 2.99 | 243.72 | 44.30 | 2.67 | (C) 1.92 | 4.01 |
| Number of Answers Learned | 79.84 | 26.90 | 1.63 | 74.28 | 25.00 | 1.50 | (C) 5.56 | 2.21 |
| Per cent of Possible Gain | 44.52 | 15.20 | .92 | 46.12 | 14.90 | .90 | (X) 1.60 | 1.28 |

the difference is 2.21 answers. On the other hand, the difference in mean per cent of possible gain is 1.60 per cent points in favor of the experimental group, but this is not a significant difference, since the probable error of the difference is 1.28 per cent points.

Table VIII, which deals with the pupils in the middle fifty per cent of the I.Q. distribution, confirms the findings for the total group. It gives additional evidence of the difference in mean number of answers learned (1.91) favoring very slightly the control group, whereas the difference in mean per cent of possible gain (2.54) favors very slightly the experimental group, but in neither case is the difference significant.

Table IX, which deals with the pupils in the upper fifteen per cent of the I.Q. distribution, further confirms the findings for the total and middle groups. It should be noted, however, that the difference of 14.21 additional answers learned favors the control group and is more than three times the probable error of the difference of the means (4.38). On the other hand, the difference in mean per cent of possible gain is 5.32 per cent points in favor of the experimental group, but this is not a significant difference, since the probable error is 3.11 per cent points.

Table X shows that for the pupils in the lower thirty-five per cent of the I.Q. distribu-

TABLE VIII

COMPARISON OF RESULTS OF PUPILS IN THE MIDDLE 50 PER CENT OF THE I.Q. DISTRIBUTION

| | Method 1 Control Group (N= 63) | | | Method 3 Exp. Group (N= 63) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Pretests | 173.73 | 34.20 | 2.90 | 171.35 | 27.90 | 2.37 | (C) 2.38 | 3.74 |
| End tests | 256.59 | 45.50 | 3.86 | 252.30 | 39.80 | 3.38 | (C) 4.29 | 5.42 |
| Number of Answers Learned | 82.86 | 29.90 | 2.54 | 80.95 | 26.40 | 2.24 | (C) 1.91 | 3.38 |
| Per Cent of Possible Gain | 45.64 | 15.70 | 1.33 | 48.18 | 14.70 | 1.25 | (X) 2.54 | 1.82 |

TABLE IX

COMPARISON OF RESULTS OF PUPILS IN THE UPPER 15 PER CENT OF THE I.Q. DISTRIBUTION

| | Method 1 Control Group (N= 19) | | | Method 3 Exp. Group (N= 19) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Pretests | 186.05 | 39.10 | 6.04 | 204.47 | 29.90 | 4.63 | (X) 18.42 | 7.61 |
| End Tests | 275.53 | 49.10 | 7.59 | 279.74 | 35.30 | 5.46 | (X) 4.21 | 9.35 |
| Number of Answers Learned | 89.48 | 20.30 | 3.14 | 75.27 | 19.80 | 3.06 | (C) 14.21 | 4.38 |
| Per Cent of Possible Gain | 46.00 | 14.50 | 2.24 | 51.32 | 14.00 | 2.16 | (X) 5.32 | 3.11 |

TABLE X

COMPARISON OF RESULTS OF PUPILS IN THE LOWER 35 PER CENT OF THE I.Q. DISTRIBUTION

| | Method 1 Control Group (N= 43) | | | Method 3 Exp. Group (N= 43) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Pretests | 145.23 | 25.90 | 2.66 | 148.26 | 30.50 | 3.14 | (X) 3.03 | 4.11 |
| End Tests | 216.40 | 39.90 | 4.10 | 219.88 | 41.20 | 4.24 | (X) 3.48 | 5.90 |
| Number of Answers Learned | 71.17 | 25.10 | 2.58 | 71.62 | 23.90 | 2.46 | (X) 0.45 | 3.56 |
| Per Cent of Possible Gain | 38.26 | 11.20 | 1.15 | 40.81 | 13.70 | 1.41 | (X) 2.55 | 1.82 |

tion there is no significant difference. However, it should be noted that in both of the measures of the success of pupils, the differences, while small (0.45 additional answers learned and 2.55 additional per cent points of possible gain), are in favor of the experimental groups.

### Summary of Significant Findings

Figure 2, which shows the percentile distributions of answers learned expressed as percentage of possible gain, confirms the general findings of the preceding tables. For instance, Figure 2 shows that for the total group neither method is definitely better. It shows that Method 1 is slightly better for the pupils in the upper part of the distribution of percentage of possible gain, whereas Method 3 is slightly better for most of the pupils in the lower part of the distribution of percentage of possible gain.

For pupils in the upper part of the dis-

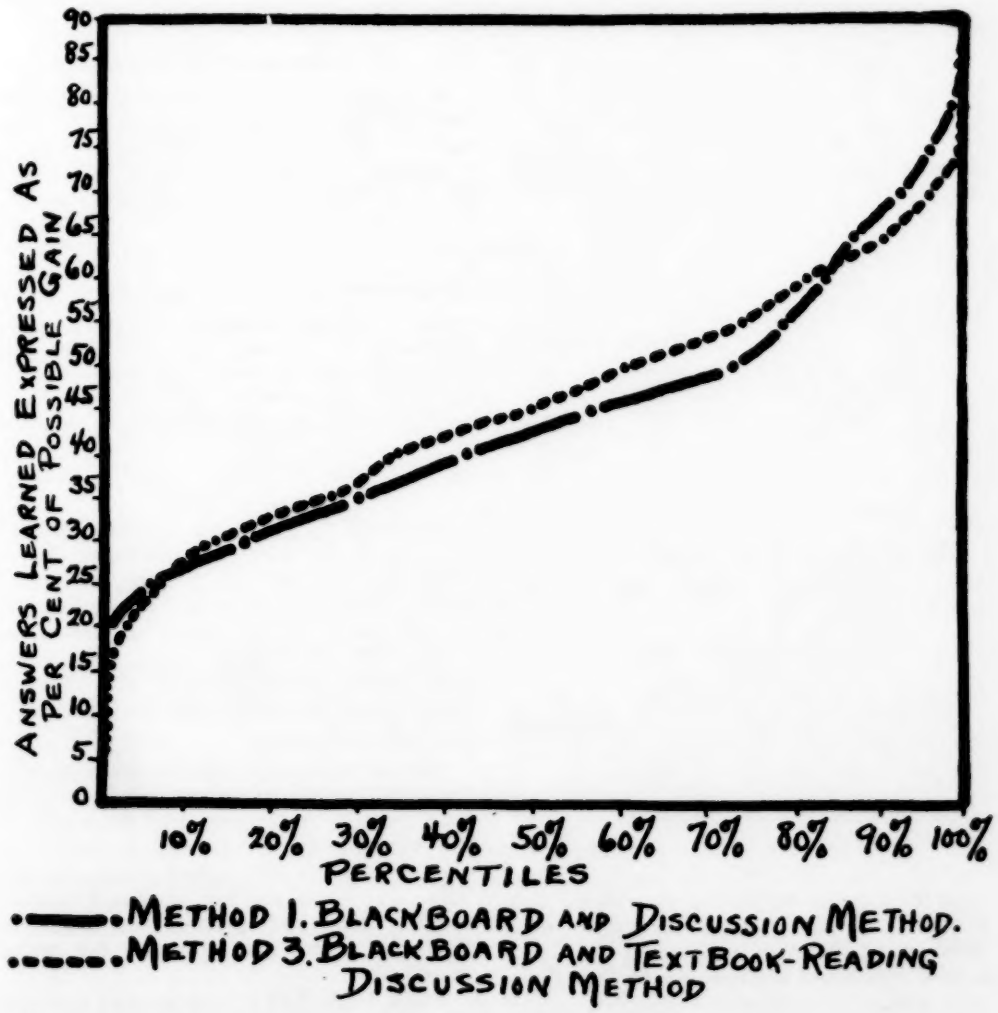



FIG. 2.-EXPERIMENT NO. 2, PERCENTILE DISTRIBUTIONS OF ANSWERS LEARNED EXPRESSED AS PER CENTS OF POSSIBLE GAIN.

tribution of percentage of possible gain, Method 1 is more successful. This is in contradiction to the results of findings in Table IX which shows that pupils in the upper part of the I.Q. distribution were more successful with Method 3. This is evidence of the unreliability of definite conclusions as to the superiority of either method for the upper I.Q.-upper achievement group.

## IV

### Experiment No. 3

*The Problem*

What are the relative differences in efficiency of Method 1 (the blackboard and discussion method), compared with Method 4 (the unit assignment method), when the methods are applied to teaching the units or topics in junior business training?

*Procedure*

The procedure is the same as in Experiment No. 1, except where differences are noted.

### Collection of Data

*Selection of Teachers and Schools*

One of the two teachers in the experiment had taught junior business training in both day junior and evening high schools.

*Time and Duration of Experiment*

Before conducting this experiment, a method somewhat similar to Method 4 had been tried by ten teachers in the evening high schools. In this method the pupils were given lesson-sheets to work on at their own rate of speed and without any time limit for each unit. This method proved unsatisfactory because in a relatively short time each pupil in the class was working on a different lesson-sheet so that class or group instruction was impossible and individual instruction was not feasible. Hence, it was decided that in this experiment a time limit should be set for the study of each unit. The number of days allotted for independent work on the unit was to be equal to the number of individual lessons in the unit. In this way, the time allotment for the study of each unit was the same for each method in this experiment.

*Data Provided by the Teachers*

A total of 264 pupil-record cards were received from the two teachers, as shown in Table XI.

### Treatment of Data

In Experiment No. 3 some of the pupils were of the low ninth (9A) grade. Their median age was approximately 14 years, 5 months. The remainder were of the high ninth (9B) grade. Their median age was approximately 14 years, 10 months. The 9A units or topics were the Intelligent Buyer, Business Procedure in Buying, Successful Selling, Selling in Business, Transportation of People, and Transportation of Goods by Mail. The 9B units or topics were Money and Credit, Thrift, Budgets and Records, Investments, Insurance, and Bank Services.

Table XI

Sources of Data for Experiment No. 3

| School No. | Grade | No. Teachers | No. Classes | Pupils, Method 1 No. | Pupils, Method 1 Mean I.Q. | Pupils, Method 4 No. | Pupils, Method 4 Mean I.Q. |
|---|---|---|---|---|---|---|---|
| 1 | 9A | 1 | 3 | 32 | 105 | 36 | 102 |
| | | | -- | -- | --- | 36 | 103 |
| *2 | 9B | 1 | 4 | 40 | 104 | 42 | 102 |
| | | | -- | 33 | 109 | 45 | 99 |
| | | 2 | 7 | 105 | 106 | 159 | 101 |

*Equating the Groups*

Further equating of the groups resulted in 184 pupil-records or seventy per cent of the total group in the experiment. Each pupil in the experimental group was matched with another pupil in the control group who had the: (1) same number of junior business training tests, both pretests and end tests; (2) same grade; and (3) same I.Q. within three points.

From Table XII it may be seen that each group was equal as to mean I.Q. and that the

* In school No. 2 all three experiments happened to be conducted; otherwise, each experiment was conducted in a different school. This Experiment No. 3 was conducted by teachers and with pupils who were not in Experiments No. 1 and No. 2.

difference of the mean of the pretests, although it was 7.39 questions in favor of the control group, was not significant. The difference was less than twice as large as the probable error of the difference (3.78).

*Summary of the Results of the Junior Business Training Tests*

The average number of questions answered by each of the 184 pupils was 694 questions, of which half were on the pretests and half on the end tests.

## Findings

Tables XII to XV summarize the results of Experiment No. 3.

Table XII shows that for the total group there is no significant difference in the means of the control and experimental groups either in the number of answers learned or in the per cent of possible gain. The mean number of answers learned by the control group or Method 1 was 66.73 and by the experimental group or Method 4 was 70.00. This difference of 3.27 additional answers learned favors the experimental group, but is not significant, since the probable error of the difference is 2.37 answers. The difference in mean percentage of possible gain is 1.63 per cent points in favor of the experimental group. Again, this is not a significant difference, since the probable error of the difference is 1.40 per cent points.

It should be noted, however, that in both of these measures of the success of pupils, the differences, while small, are in favor of the experimental groups.

Table XIII deals with the pupils in the middle fifty-seven per cent of the I.Q. distribution. It confirms the findings for the total group in that it shows the differences in the mean number of answers learned (4.62) and the mean per cent of possible gain (1.54) are not significant, although both differences favor slightly the control groups.

Table XIV concerns the pupils in the upper sixteen per cent of the I.Q. distribution. It shows the difference in mean number of answers learned to be 0.67 answers in favor of the experimental group and the difference in mean per cent of possible gain to be 2.00 per cent points in favor of the same group, but in neither case is the difference significant according to the probable error of the difference.

Table XV deals with the pupils in the lower twenty-seven per cent of the I.Q. distribution. It shows a difference of 18.00

TABLE XII

COMPARISON OF RESULTS FOR ALL PUPILS

| | Method 1 Control Group (N= 92) | | | Method 4 Exp. Group (N= 92) | | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | | |
| Phila. Mental Ability Test I. Q. | 105.4 | 9.17 | 0.64 | 105.4 | 9.55 | 0.67 | | 0.0 | 0.92 |
| Pretests | 203.48 | 38.20 | 2.68 | 196.09 | 39.60 | 2.67 | (C) | 7.39 | 3.78 |
| End Tests | 270.21 | 46.90 | 3.30 | 266.09 | 57.90 | 4.07 | (C) | 4.12 | 5.24 |
| Number of Answers Learned | 66.73 | 18.50 | 1.30 | 70.00 | 28.20 | 1.98 | (X) | 3.27 | 2.37 |
| Per Cent of Possible Gain | 43.15 | 12.20 | .86 | 44.78 | 15.80 | 1.11 | (X) | 1.63 | 1.40 |

TABLE XIII

COMPARISON OF RESULTS OF PUPILS IN THE MIDDLE 57 PER CENT OF THE I.Q. DISTRIBUTION

| | Method 1 Control Group (N= 52) | | | Method 4 Exp. Group (N= 52) | | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | | |
| Pretests | 203.27 | 35.40 | 3.31 | 190.19 | 40.50 | 3.79 | (C) | 13.08 | 5.03 |
| End tests | 272.89 | 39.90 | 3.73 | 255.19 | 60.70 | 5.68 | (C) | 17.70 | 6.79 |
| Number of Answers Learned | 69.62 | 16.80 | 1.57 | 65.00 | 27.90 | 2.61 | (C) | 4.62 | 3.04 |
| Per Cent of Possible Gain | 44.62 | 9.60 | .90 | 43.08 | 15.40 | 1.44 | (C) | 1.54 | 1.70 |

TABLE XIV

COMPARISON OF RESULTS OF PUPILS IN THE UPPER 16 PER CENT OF THE I.Q. DISTRIBUTION

| | Method 1 Control Group (N= 15) | | | Method 4 Exp. Group (N= 15) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Pretests | 220.33 | 31.40 | 5.47 | 216.33 | 31.80 | 5.54 (C) | 4.00 | 7.78 |
| End Tests | 298.33 | 29.60 | 5.16 | 295.00 | 46.70 | 8.14 (C) | 3.33 | 9.64 |
| Number of Answers Learned | 78.00 | 10.90 | 1.90 | 78.67 | 27.50 | 4.80 (X) | .67 | 5.16 |
| Per Cent of Possible Gain | 50.33 | 10.90 | 1.89 | 52.33 | 14.80 | 2.58 (X) | 2.00 | 3.20 |

TABLE XV

COMPARISON OF RESULTS OF PUPILS IN THE LOWER 27 PER CENT OF THE I.Q. DISTRIBUTION

| | Method 1 Control Group (N= 25) | | | Method 4 Exp. Group (N= 25) | | | Dif. of Means | P. E. Dif. of Means |
|---|---|---|---|---|---|---|---|---|
| | Mean | S. D. | P. E. Means | Mean | S. D. | P. E. Means | | |
| Pretests | 193.80 | 43.50 | 5.87 | 196.20 | 37.80 | 5.10 (X) | 2.40 | 7.78 |
| End tests | 247.80 | 57.10 | 7.70 | 268.20 | 51.50 | 6.95 (X) | 20.40 | 10.37 |
| Number of Answers Learned | 54.00 | 22.70 | 3.06 | 72.00 | 27.30 | 3.68 (X) | 18.00 | 4.78 |
| Per Cent of Possible Gain | 35.40 | 14.10 | 1.90 | 43.80 | 16.10 | 2.17 (X) | 8.40 | 2.88 |

answers in the mean number of answers learned in favor of the experimental group and the difference is practically four times greater than the probable error of the difference (4.78). It also shows 8.40 per cent points difference in the mean percentage of possible gain in favor of the experimental group, and the difference is nearly three times greater than the probable error of the difference (2.88).

### SUMMARY OF SIGNIFICANT FINDINGS

Figure 3 shows the percentile distributions of answers learned expressed as percentage of possible gain. These distributions confirm the findings shown in Tables XII, XIII, and XIV for the total, middle, and upper I.Q. groups, but conflict with the findings for the lower I.Q. groups (Table XV). Both Figure 3 and Table XIV show slight advantages of Method 4 for the upper part of the distribution of percentage of possible gain and the upper part of the I.Q. distribution, respectively. However, Figure 3 shows a slight advantage of Method 1 for the lower part of the distribution of percentage of possible gain in contrast to a slight advantage of Method 4 for the lower part of the I.Q. distribution (Table XV). This is evidence of the unreliability of definite conclusions as to the superiority of either method for the lower I.Q.-lower achievement group.

## V

### GENERAL SIGNIFICANCE OF THE FINDINGS

In a series of three experiments, three varying methods of teaching junior business training were compared with one control method. Parallel or equivalent groups of pupils were used to determine the superiority, if any, of any one of the four methods. Results were studied for each of the total groups and for subdivisions of the total groups according to their relative achievement, whether high, average, or low.

In no case has clear-cut evidence been found as to the superiority of any one method at any point in the achievement range. Accordingly, the selection of the method to be used should depend upon other factors than the ability and achievement of the pupils. Such factors are:

(1) available equipment, including lesson-sheets.
(2) the teacher's training and personality.
(3) the learning habits of the pupils.

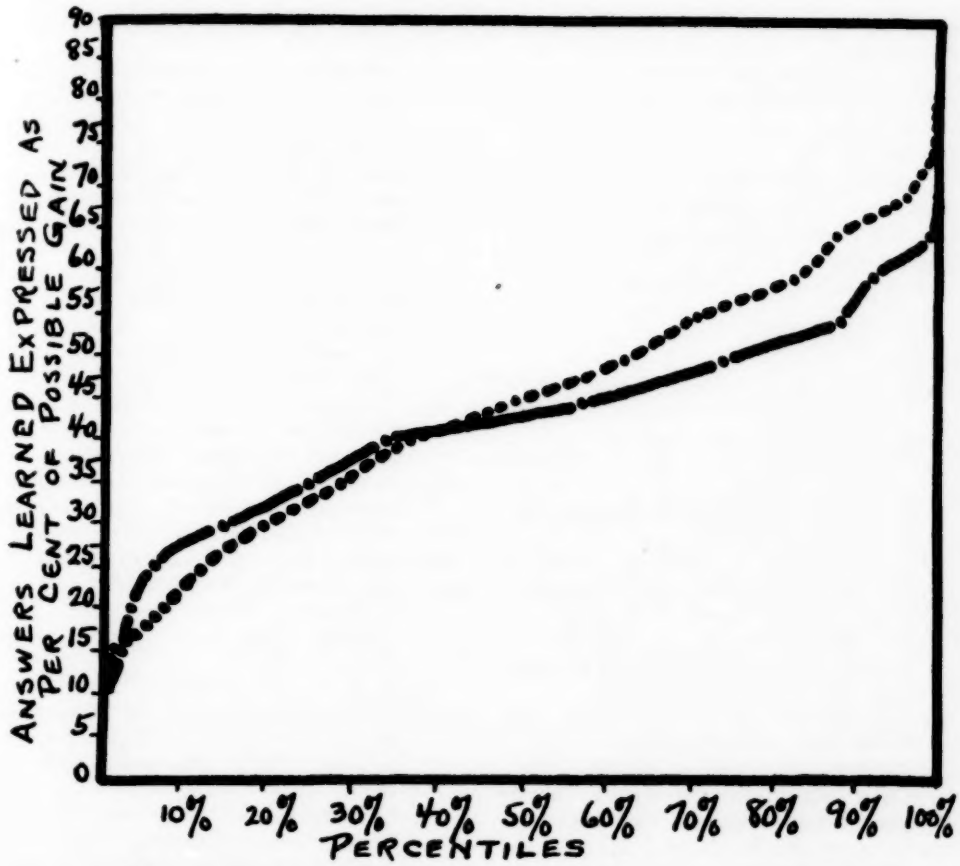



FIG. 3.—EXPERIMENT NO. 3, PERCENTILE DISTRIBUTIONS OF ANSWERS LEARNED EXPRESSED AS PER CENTS OF POSSIBLE GAIN.

However, there is slight evidence that the high ability pupils profit by the greater freedom offered them with prepared lesson-sheets (Methods 2 and 4). Pupils in the low ability groups do not seem to profit by the use of lesson-sheets. However, the evidence on this point is conflicting in Experiment No. 3 where the lesson-sheets were used for complete unit assignments.

The factor of reading in class was contrasted with assigned reading at home in Experiment No. 2. Reading in class seems to favor the low ability groups, while use of class time for other purposes seems to favor the high ability groups. On this last point, the evidence is again conflicting, depending on how the high ability group is defined. Thus, the selection of the high ability group as the upper fifteen per cent of the I.Q. distribution shows an advantage for textbook reading for this group (Table IX). On the other hand, the selection of the upper fifteen per cent of the group according to percentage of possible gain shows an advantage for the

method not requiring textbook reading in class (Figure 2). Discrepancies such as these may be due to the relatively small numbers of pupils involved when the total groups were subdivided into high-, middle-, and low-ability groups.

To determine the relative success of the various methods, it would be desirable to study the achievement of pupils actually segregated into ability groups large enough to furnish results which would be reliable.

## Bibliography

*A. General*


1. Almack, J. C. *Research and Thesis Writing.* Boston: Houghton–Mifflin Co., 1930. Pp. xvii + 310.
2. Blackstone, E. G. "Special Methods on High School Level, Commercial Subjects," *Review of Educational Research,* II (February, 1932), 67–70.
3. Blackstone, E. G. (Editor). *Research Studies in Commercial Education.* University of Iowa Monographs in Education, Vols. I, II, III, IV, V. Iowa City, Iowa: University of Iowa, 1926, 1928, 1928, 1929, 1932.
4. Buckley, H. B. "Supervision and Helps for Teachers Applied to Junior Business Training in Philadelphia." Unpublished Master's thesis, Temple University, Teachers College, 1928.
5. Cheyney, W. W. *Rate of Progress and I.Q.* Philadelphia, Pa.: University of Pennsylvania, 1930. Pp. 98.
6. Drake, Richard. "A Comparison of Two Methods of Teaching High School Algebra," *Journal of Educational Research,* XXIX (September, 1935), 12–16.
7. Eastern Commercial Teachers Association. *Modern Methods of Teaching Business Subjects.* Philadelphia, Pa.: Banks College—Fourth Yearbook, 1931, and Fifth Yearbook, 1932.
8. Garrett, H. E. *Statistics in Psychology and Education.* New York: Longmans, Green and Co., 1936. Pp. xiii + 317.
9. Good, C. V. *How to Do Research in Education.* Baltimore: Warwick and York, Inc., 1928. Pp. 298.
10. Gordon, H. C. *The Specific Nature of Achievement and the Predictive Value of the I.Q.* Philadelphia, Pa.: University of Pennsylvania, 1931. Pp. 147.
11. Gropp, Regina. "A Guide to the Current Literature in Business Education," *Journal of Business Education,* X (June, 1935), 19–20.
12. Haas, K. B. "A Bibliography on Consumer Education for Business," *National Business Education Quarterly,* II (May, 1934), 24–26, 30.
13. Haynes, B. R., and Graham, J. *Research in Business Education.* Los Angeles, California: C. C. Crawford, University of Southern California, 1932. Pp. 232.
14. Holzinger, K. J. *Statistical Methods for Students of Education.* New York: Ginn and Co., 1928. Pp. 372.
15. Kandel, I. L. *Twenty-Five Years of American Education,* pp. 141–76. New York: Macmillan Co., 1926.
16. Kirk, John G., Buckley, Harold B., and Waesche, Mary A. *Introduction to Business, Manual and Key.* Philadelphia: John C. Winston Co., 1933. Pp. vi + 250.
17. N. E. A., Department of Superintendence. *Junior Commercial Curriculum, Specific Studies and References,* pp. 445–54. Fifth Yearbook, February, 1927.
18. Nichols, F. G. "Looking Ahead in Junior Business Training Instruction," *National Business Education Quarterly,* I (October, 1932), 20–28.
19. Odell, W. R. "Research in Commercial Education," *Journal of Business Education,* X (April, 1935), 11.
20. Reeder, W. G. *How to Write a Thesis.* Bloomington, Illinois: Public School Publishing Co., 1930. Pp. x + 216.
21. Shear, B. E. "Getting Results in Business Training," *Journal of Business Education,* VIII (September, 1932), 15–16.
22. Seyfried, J. E. *Principles and Mechanics of Research.* University of New Mexico Bulletin, Vol. IX, No. 1. Albuquerque, New Mexico: University of New Mexico Press, 1935. Pp. 240.
23. Trow, W. C. *Scientific Method in Education.* Boston: Houghton–Mifflin Co., 1925. Pp. xi + 159.

24. U. S. Bureau of Education. *An Outline of Methods of Research.* Bulletin No. 24, 1926. Washington, D. C.: U. S. Government Printing Office, 1926. Pp. 31.
25. U. S. Bureau of Education. *Commercial Education, Junior Business Training.* Bulletin No. 26050. Washington, D. C.: U. S. Government Printing Office, August, 1928.
26. U. S. Bureau of Education. *Commercial Education, Tests and Measurements.* Bulletin No. 22226. Washington, D. C.: U. S. Government Printing Office, March, 1928.
27. U. S. Bureau of Education. *Junior Business Training.* Bulletin No. 61768. Washington, D. C.: U. S. Government Printing Office, March 22, 1932.
28. Walters, R. G. *Modern Methods of Teaching Commercial Subjects.* Monograph No. 16. Cincinnati, Ohio: South-Western Publishing Co., 1933. Pp. 31.

*B. Recent Studies Relating to Junior Business Training*

1. Anderson, Ray L. "Junior Business Training in the Public Secondary Schools of the Pacific Coast." Unpublished Master's thesis, University of Southern California, 1935.
2. Berry, Ralph Edward. "A Study of the Work of Juniors in Telegraph Offices with Suggestions for Training." Unpublished Master's thesis, University of California.
3. Blum, Hazel. "A Program of Testing in Junior Business Training." Unpublished Master's thesis, University of Pittsburgh, 1933.
4. Boggs, Rose E. "An Analysis of Textbooks and Courses of Study to Determine Course-of-Study Content in Junior Business Training." Unpublished Master's thesis, Indiana University.
5. Briscoe, Laura. "A Study of the Relative Effectiveness of Teaching Penmanship in Separate Classes and in Connection with Junior Business Training." Unpublished Master's thesis, University of Iowa, 1935.
6. Brown, Harry A. "Enriched Material for Junior Business Training." Unpublished Master's thesis, University of Iowa.
7. Burchfield, Howard E. "Most Common Business Mistakes Made by Laymen." Unpublished Master's thesis, Greeley State Teachers College, 1934.
8. Cardy, Earl N. "Recent Trends and Developments in the Commercial Curriculum of the Junior High School." Unpublished Master's thesis, Boston University, 1935.
9. Davis, N. Evelyn. "A Study of Junior Business Training Needs in George Washington Junior High School." Unpublished Master's thesis, Stanford University, 1928.
10. Fasnacht, Harold D. "An Evaluation of Junior Business Training." Unpublished Master's thesis, Greeley State Teachers College.
11. Fisk, William McKee. "The Relation Between Business Arithmetic and Business Practice." Unpublished Master's thesis, University of Southern California, 1926.
12. Gibbons, Francis F. "The Development of Junior Business Training in the High School." Unpublished Master's thesis, Boston University, 1935.
13. Gilman, Grace H. "An Investigation of the Use of Dramatized Materials as a Teaching Device in Junior Business Training." Unpublished Master's thesis, University of Southern California.
14. Hanson, Martha. "Business Skills and Information Needed by Everyone as Determined by an Investigation of Experiences of Laymen in North Dakota." Unpublished Master's thesis, Greeley State Teachers College.
15. Hess, Edmund N. "A Study of the Junior Business Activities of the Pupils of Barratt Junior High School." Unpublished Master's thesis, Temple University, 1935.
16. Levine, Harry M. "A Comparison of General Business Training Outside of Philadelphia with That in Junior High Schools of Philadelphia." Unpublished Master's thesis, Temple University, 1935.
17. Ritter, Harvey R. "An Investigation of the Content of a Course in Personal

Economics for the Ninth Grade of the Junior High School." Unpublished Master's thesis, University of Southern California, 1928.

18. Salmond, M. G. "Junior Business Training." Unpublished Master's thesis, Boston University.
19. Sevin, Rose. "A Study of the Socio-Economic Background and a Survey of the Interests of Ninth-Grade Commercial Pupils in the Herron Hill Junior High School." Unpublished Master's thesis, University of Pittsburgh, 1934.
20. Shaw, Ray Banker. "Non-Vocational Aims of Commercial Education in Secondary Schools." Unpublished Master's thesis, University of Southern California, 1930.
21. Warren, Dorothy E. "A Study to Determine the Business Practices Experienced by Intermediate Grade Pupils." Unpublished Master's thesis, University of Southern California.
22. Warren, D. E., and Burton, W. H. "Knowledge of Simple Business Practices Possessed by Intermediate Grade Pupils," *Elementary School Journal*, XXXV (March, 1935), 511–16.
23. Werlinsky, Daniel. "A Questionnaire Study of Business Activities Performed by Junior High School Pupils." Unpublished Master's thesis, Temple University, 1933.

## *Appendix*

TYPEWRITTEN SHEET OF DIRECTIONS FOR EACH TEACHER CONCERNING METHODS OF TEACHING IN EXPERIMENT NO. 1

**Directions for Comparing**

**Method 1, the blackboard and discussion method, with**
**Method 2, the lesson-sheet and textbook-reading discussion method**

***Procedure***

| Method 1 | Method 2 |
|---|---|
| I. *Self-reliance work:* | |
| This part of the lesson should be the same for each class as to content, methods, and time-allotment (13 minutes) except as follows: | |
| Method 1 class (the control group) will copy the work from the blackboard. | Method 2 class (the experimental group) will copy the work from the mimeographed lesson-sheet. |
| II. *Problem:* | |
| This part of the lesson should be the same for each class as to content, methods, and time-allotment (13 minutes) except as follows: | |
| Problem of the lesson to be solved by: | |
| 1st—Class discussion led by teacher.<br>2nd—Development of an outline, diagram, or model business form by the teacher at the blackboard. | 1st—Silent textbook reading of pages related to lesson-problem. Maximum time allowed for reading to be 5 minutes.<br>2nd—Class discussion led by teacher.<br>3rd—Filling in incomplete statements on lesson-sheet. |
| III. *Practice work:* | |
| This part of the lesson should be the same for each class as to content, methods, and time-allotment (13 minutes) except as follows: | |
| Method 1 class will copy practice work from blackboard. | Method 2 class will copy practice work from lesson-sheet. |
| IV. *Homework:* | |
| This should be the same for each class. | Pupils in Method 2 class will keep the lesson-sheets in their notebooks. |

*Mimeographed Sheet for Reporting Class Results in Experiment No. 1*

SCHOOL DISTRICT OF PHILADELPHIA, DEPARTMENT OF SUPERINTENDENCE
DIVISION OF COMMERCIAL EDUCATION

REPORT OF PUPILS' ACHIEVEMENT ON TESTS[1]

Section---------- Used Method No.------ Teacher ----------

School ----------

| Pupils Alphabetical Order | I. Q. | Test No. 1* R* | P | A | Test No. 2 R | P | A | Test No. 3 R | P | A | Test No. 4 R | P | A | Test No. 5 R | P | A | Test No. 6 R | P | A | Test No. 7 R | P | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. ** | | | | | | | | | | | | | | | | | | | | | | |
| 2. | | | | | | | | | | | | | | | | | | | | | | |
| 3. | | | | | | | | | | | | | | | | | | | | | | |
| 58. | | | | | | | | | | | | | | | | | | | | | | |
| 59. | | | | | | | | | | | | | | | | | | | | | | |
| Class Mean | | | | | | | | | | | | | | | | | | | | | | |
| Class Median | | | | | | | | | | | | | | | | | | | | | | |

CODE:
*R = Pupils raw score for review test
P = Pretest mark
A = (R—P = A) Answers Learned
**—Encircle numbers indicating boys. Write last name only.

*Test No. 1—----------(----------)----
(Name of Test) (Time)
Test No. 2—----------(----------)----
Test No. 3—----------(----------)----
Test No. 4—----------(----------)----
Test No. 5—----------(----------)----
Test No. 6—----------(----------)----
Test No. 7—----------(----------)----

(Time allowed should be the same for each pretest and review test; i. e., 20 min. for term test and 15 min. for topic tests.)

(1) Term pretest and review test contained 56 questions: Nos. 1-50 and 95-100 of Term Pretest and Review Pretest No. 1 of Introduction to Business Pretests and Review Tests by Kirk, Buckley, and Waesche (Published by The John C. Winston Co., Phila.)

Tests No. 1, 2, 3, 4, 5, and 6 respectively were unit or topic tests of fifty questions each on the Business Letter; The Telephone; Telegrams, Cablegrams, and Radiograms; Getting Along with People; Sources of Business Information; and Filing and Finding Information. They are called Topic Pretests and Review Tests because they are used as both pretests and as review or end tests.

REPORT CARD OF INDIVIDUAL PUPIL'S RECORD IN EXPERIMENT NO. 1

L.N. Canning
P.Oc. Secretary

| Term | R.= 51 | P.= 36 | A.= 15 | A.P.= 20 |
|---|---|---|---|---|
| B.L. | 41 | 33 | 11 | 17 |
| Tp. | 46 | 22 | 24 | 28 |
| Tg. | 44 | 25 | 19 | 25 |
| C.P. | 49 | 43 | 6 | 7 |
| B.I. | 48 | 40 | 8 | 10 |
| F. | 42 | 30 | 12 | 20 |
| Total | 324 | 229 | 95 | 127 |

$\frac{A}{A.P.} = \frac{95}{127}$ = 74% of Possible Gain

F. Age (Y) 13 (M) 6 C 1
F. R.S. 17 I.Q. 128
Ab. 0 days
Ab. 0
Ab. 0
Ab. 0
Ab. 0
Ab. 0
Ab. 0

| Report | Marks |
|---|---|
| (1) | E |
| (2) | E+ |
| (3) | E+ |

School Wg.
Repeater No

L.N.—*Last name;* M or F—*male or female;* Age (Y) *years* (M) *months;* C or X—in control or experimental class; 1, 2, 3—in Experiment No. 1, 2, or 3.

P.Oc.—Parent's occupation; F or M—father or mother; R.S.*—reading score; I.Q.**—intelligence quotient.

Term: R = term review test, number of answers correct; P = term pretest, number of answers correct; A = achievement or Number of Answers Learned for term test, shown as the difference between the pretest score and the review test score; A.P. = number of Answers Possible to learn; Ab—number of days absent during the term.

B.L.—Business letter test results: R—review test; P—pretest; A—answers learned or difference between R and P.

Tp.—Telephone topic tests results: R—review test; P—pretest; A—answers learned or difference between R and P.

Tg.—Telegraph topic tests results: R—review test; P—pretest; A—answers learned or difference between R and P.

C.P.—Communicating in person topic results: R—review test; P—pretest; A—answers learned or difference between R and P.

B.I.—Business information topic results: R—review test; P—pretest; A—answers learned or difference between R and P.

F.—Filing and Finding information results: R—review test; P—pretest; A—answers learned or difference between R and P.

Report marks: (1) for 1st report period, at end of 6 weeks.
(2) for 2nd report period, at end of 12 weeks.
(3) for 3rd report period, at end of 18 weeks.

School (*initial* of school to show from which school the card came).
Repeater? ------ Yes or No, as to whether or not pupil is repeating the grade.
A ÷ A.P. = Per Cent of Possible Gain.

* R.S.—reading score obtained by teacher after using Directions for Administration and Scoring of Tests issued by Division of Educational Research and Results, Philadelphia, Pa., and known as Philadelphia Silent Reading Test for Comprehension and Rate.

** I.Q.—intelligence quotient obtained by teacher from results of Philadelphia Mental Ability Test for 8B issued by Division of Educational Research and Results.